Evaluating DNA Barcoding for Species Identification and Discovery in European Gracillariid Moths

Carlos Lopez-Vaamonde1,2,*, Natalia Kirichenko3,4, Alain Cama5, Camiel Doorenweerd6,7, H. Charles J. Godfray8, Antoine Guiguet9, Stanislav Gomboc3, Peter Huemer11, Jean-François Landry12, Ales Laštůvka13, Zdenek Laštůvka14, Kyung Min Lee15, David C. Lees16, Marko Mutanen15, Erik J. van Nieukerken7, Andreas H. Segerer17, Paolo Triberti18, Christian Wieser19 and Rodolphe Rougerie20

1 INRAE, URZF, Orléans, France, 2 IRBI, UMR 7261, CNRS-Université de Tours, Tours, France, 3 Sukachev Institute of Forest, Siberian Branch of Russian Academy of Sciences, Federal Research Center “Krasnoyarsk Science Center SB RAS,” Krasnoyarsk, Russia, 4 Siberian Federal University, Krasnoyarsk, Russia, 5 La Chapelle sur Loire, France, 6 Department of Plant and Environmental Protection Sciences, College of Tropical Agriculture and Human Resources, University of Hawaii, Honolulu, HI, United States, 7 Naturals Biodiversity Center, Leiden, Netherlands, 8 Department of Zoology, University of Oxford, Oxford, United Kingdom, 9 Department of Biology, The Pennsylvania State University, University Park, PA, United States, 10 Gancani, Beltinci, Slovenia, 11 Tiroler Landesmuseum Betriebsges.m.b.H., Sammlungsm- und Forschungszentrum, Hall in Tirol, Austria, 12 Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON, Canada, 13 Prostějov, Czechia, 14 Faculty of Agriculture, Mendel University, Brno, Czechia, 15 Ecology and Genetics Research Unit, University of Oulu, Oulu, Finland, 16 Department of Life Sciences, Natural History Museum, London, United Kingdom, 17 Staatliche Naturwissenschaftliche Sammlungen Bayerns, Zoologische Staatsammlung München (SNSB-ZSM), Munich, Germany, 18 Museo Civico di Storia Naturale, Zoologia, Verona, Italy, 19 Landesmuseum für Kärnten, Abteilung Zoologie, Sammlungs- und Wissenschaftszentrum, Klagenfurt, Austria, 20 Institut de Systématique, Evolution, Biodiversité (ISYE), Muséum national d’Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

Gracillariidae is the most species-rich leaf-mining moth family with over 2,000 described species worldwide. In Europe, there are 263 valid named species recognized, many of which are difficult to identify using morphology only. Here we explore the use of DNA barcodes as a tool for identification and species discovery in European gracillariids. We present a barcode library including 6,791 COI sequences representing 242 of the 263 (92%) resident species. Our results indicate high congruence between morphology and barcodes with 91.3% (221/242) of European species forming monophyletic clades that can be identified accurately using barcodes alone. The remaining 8.7% represent cases of non-monophyly making their identification uncertain using barcodes. Species discrimination based on the Barcode Index Number system (BIN) was successful for 93% of species with 7% of species sharing BINs. We discovered as many as 21 undescribed candidate species, of which six were confirmed from an integrative approach; the other 15 require additional material and study to confirm preliminary evidence. Most of these new candidate species are found in mountainous regions of Mediterranean countries, the South-Eastern Alps and the Balkans, with nine candidate species found only on islands. In addition, 13 species were classified as deep conspecific lineages, comprising a total of 27 BINs with no intraspecific morphological differences found, and no known ecological differentiation. Double-digest restriction-site associated DNA sequencing (ddRAD) analysis showed strong mitonuclear discrepancy.
in four out of five species studied. This discrepancy is not explained by Wolbachia-mediated genetic sweeps. Finally, 26 species were classified as “unassessed species splits” containing 71 BINs and some involving geographical isolation or ecological specialization that will require further study to test whether they represent new cryptic species.

Keywords: barcode index number, cryptic diversity, Gracillariidae, Lepidoptera, mitonuclear discordance, new candidate species, RAD sequencing, Wolbachia

INTRODUCTION

Alarming projections of biodiversity loss indicate that we have entered the world’s sixth mass extinction (Raven and Miller, 2020). Meanwhile, only a small fraction of species on Earth have been formally described and assigned a scientific name (Scheffers et al., 2012), so many taxa are likely to go extinct before we know of their existence (Eggleton, 2020). The incompleteness of our current census of life, referred as the “Linnean shortfall” (Raven and Wilson, 1992), is especially acute for the most diverse yet least studied groups such as small insects, and for the high diversity tropics (Delaby et al., 2019; Lopez-Vaamonde et al., 2019).

The increasing use of genetic tools in species discovery has shown the shortfall to be greater than previously recognized because of the frequent discovery of sets of two or more cryptic species that have been incorrectly classified as a single taxon due to their morphological similarity. In particular, DNA barcoding, a tool for species identification based on the use of a single standard DNA marker (a fragment of the COI mtDNA gene; Hebert et al., 2003), has contributed to the discovery of many new species, including among them the most closely studied and best documented groups of organisms (Hrbek et al., 2014). Some of those new species are cryptic, rare and of conservation value; assessing the level of cryptic diversity and identifying their biodiversity hotspots should then be a conservation priority (Bickford et al., 2007).

The European Lepidoptera is one of the best-studied faunas in the world, thanks to more than two and a half centuries’ work by a large community of both professional and amateur lepidopterists. The fauna has been the focus of intensive barcoding activity with 7,831 species barcoded out of the 10,723 (73%) species of Lepidoptera known to occur in Europe (Rennwald and Rodeland, 2020) (Barcode of Life Datasystems, www.boldsystems.org; as of mid-October 2020; Ratnasingham and Hebert, 2007). Many European Lepidoptera species show deeply divergent mitochondrial intraspecific lineages raising the possibility of high levels of cryptic diversity (Dincă et al., 2011, 2013, 2015; Hausmann et al., 2013; Mutanen et al., 2013, 2016; Vodă et al., 2015; Hernández-Roldán et al., 2016). Little is known about the extent to which these barcode splits reflect biological species. A number of studies have addressed this question by looking at the correspondence between COI clusters and nuclear lineages using genome-wide DNA analysis. Some have found a significant association suggesting the presence of cryptic species (Janzen et al., 2017; Kozlov et al., 2017; Dincă et al., 2019) whereas others have found a strong discrepancy between mitochondrial and nuclear DNA (Hinojosa et al., 2019). One explanation for the discrepancy is mitochondrial drive causing the replacement of one mitotype by another due to the introduction of the intracellular bacterium Wolbachia which spreads by causing cytoplasmic incompatibility, or by another agent with the same or similar genetic strategy. Divergent COI clusters have been found to have different Wolbachia strains in some (Kodandaramaiah et al., 2013) but not all cases (Bereczki et al., 2015).

Here, we explore the diversity of European Gracillariidae, the largest family of leaf-mining micro-moths, by analyzing DNA barcode data in the context of the current taxonomy of the family on a continental scale. The Gracillariidae are one of the most species-rich families of Lepidoptera with 1996 described species worldwide, data on September 23, 2020 (De Prins and De Prins, 2020), though many more species await description, particularly in the neotropics (Lees et al., 2014). In contrast, the faunistics, alpha taxonomy and ecology of this family in Europe are relatively well known (Doorenweerd et al., 2014), with 24 genera and 263 species recorded (Supplementary Table S1). The majority of gracillariids are leaf miners (their larvae feed and live within a single leaf) and feed on a huge diversity of different host plants (De Prins and De Prins, 2020). However, most species are monophagous or oligophagous (Lopez-Vaamonde et al., 2003). Several species are serious pests of agricultural and ornamental plants (Lopez-Vaamonde et al., 2010), and some are highly invasive (Valade et al., 2009; Lees et al., 2011). Ongoing DNA barcoding campaigns represent a potential valuable source for documenting intraspecific variation and can answer questions about the phyllogeography and invasion genetics of gracillariid moths. Examples of such studies include work on European species that have expanded their distribution range like the Horse-Chestnut Leaf Miner, Phyllonorycter issikii (Kumata) (Kirichenko et al., 2017).

The striking coloration of the forewings of many species and the ease with which they can be reared, make this family an attractive group for amateur microlepidopterists to study (Emmet et al., 1985; Bengtsson and Johansson, 2011; Nel and Varenne, 2014). New gracillariid species are still being described in southern Europe (Corley, 2014; Kirichenko et al., 2015; Huemer et al., 2016; Varenne and Nel, 2016) and even in well-known areas like the UK (Langmaid and Corley, 2007).

Some species such as representatives of genus Phyllocnistis, and Fabaceae-feeding Phyllonorycter species, are notoriously difficult to identify using external morphology only or even after genitalia examination, but DNA barcoding has been shown to be
a useful tool to identify those difficult groups (Laštůvka et al., 2013; Kirichenko et al., 2018).

The main aims of this study are: (1) to develop a comprehensive DNA barcode reference library for the Gracillariidae of Europe based on morphological analysis by taxonomic experts that will facilitate reliable species identification; (2) to highlight deeply diverged mitochondrial lineages within species which represent potential cases of cryptic species; (3) to investigate deep COI splits in a selection of species by using genome-wide double-digest restriction-site associated (ddRAD) markers and Wolbachia screening; (4) to assess the number of identified but taxonomically undescribed new candidate species, their geographical distribution and host plant use.

MATERIALS AND METHODS

Species Checklist
We compiled a revised checklist of the 263 European Gracillariidae species (checklist CL-EUGRA in BOLD) following the higher classification of Kawahara et al., 2017 (Supplementary Table S1).

Our study countries encompass the 27-member states of EU as well as Albania, Bosnia and Herzegovina, Macedonia, Montenegro, Belarus, Ukraine, Russia (West of Urals), Liechtenstein, Norway, Switzerland, and United Kingdom (Figure 1), although we lacked records for a few countries (e.g., Estonia, Ireland). We refer to Europe here in a loose sense to include all these countries (Turkey, Russia east of Urals and overseas territories not included). We included five species [Parornix kumatai Ermolaev, Ph. macrantherella (Kuznetzov), Ph. populii (Filipjev), Ph. populaialbae (Kuznetzov), and Ph. laurocerasi (Kuznetzov)] that are known to occur in the northeast Caucasus (Chechnya and Dagestan), a region sometimes excluded from Europe for practical reasons (i.e., in Fauna Europaea; de Jong et al., 2014).

Taxon Sampling
Our taxon sampling strategy aimed to obtain DNA barcodes from as many different European species as possible, but also to sample multiple specimens from as many different geographic localities as possible to document genetic diversity over the European distribution range of each species. We built a DNA barcode dataset of 6,791 specimens of European Gracillariidae based on 5,383 adult specimens and 1,408 immature stages deposited in a total of 60 research collections (24 public and 36 private reference collections) (Supplementary Table S2). We used ESRI ArcGIS Pro software to map all georeferenced records from 1,840 different localities ranging from 28° to 70°N latitude; 18°W to 48°E longitude and elevations ranging from 0 to 2,300 m to show the sample coverage in European countries (Figure 1). For the base map, ESRI topographic map layers from ArcGIS online version 10.6 were used.

We preferentially barcoded recently collected individuals with 75% of specimens sequenced being between four and eight years old at time of sequencing.

Voucher specimens came from rearings (1,646 individuals), Malaise traps (1,873 individuals), light traps (1,764 individuals), or were hand-netted (379 individuals); 1,408 larvae or pupae were dissected from leaf mines. A total of 2,368 records in our dataset include food plant information from reared individuals (see Lopez-Vaamonde et al., 2021 for collecting and rearing protocols) or from immature stages dissected out of their leaf mines (Supplementary Table S2). Identification of immature stages was based on a combination of host plant data and DNA barcoding results.

Adult specimens were identified based on examination by expert taxonomists of wing pattern and in some cases genitalia (Supplementary Table S2).

Laboratory Procedures and Analyses of DNA Barcoding Data
DNA barcodes were obtained from adult specimens by removing either one hind leg or the whole abdomen for DNA extraction (in the latter case the abdomen could be recovered for genitalia analysis). For immature stages, whole bodies of larvae or pupae were used, usually stored in 96% ethanol, except in some cases where larvae were dissected from dried herbarium specimens with leaf mines.

In most cases, tissue samples were shipped to the Canadian Centre for DNA Barcoding (CCDB, Biodiversity Institute of Ontario, University of Guelph) for sequencing; other samples were processed at Naturalis Biodiversity Center (Leiden, Netherlands) or at laboratories of the authors’ research organizations. For the CCDB laboratory protocol see deWaard et al. (2008); for the Naturalis protocol see Doorenweerd et al. (2017). Primers LCO1490 and HCO2198 were used to sequence the COI barcode fragment (Folmer et al., 1994).

In total, 4,885 COI DNA barcodes (72%) were obtained with Sanger sequencing; the remaining 1,873 sequences of Malaise trap samples were obtained using Single Molecule Real-Time (SMRT) sequencing through the Sequel (PacBio) pipeline at CCDB (Hebert et al., 2018). In addition, 33 sequences were obtained from historical specimens of the Jäckh Collection deposited at the National Museum of Natural History, Smithsonian Institution (Washington DC, United States) using the protocol and workflow described in Prosser et al. (2016). Our data set also includes 217 reared barcoded adult specimens of Deschka’s major reference collection (recently deposited at ZSM, Munich). Sequences over 300 bp are assigned automatically to BINs in BOLD (Ratnasingham and Hebert, 2013).

We used PyCOIStats v1.3 (Doorenweerd et al., 2020) to calculate the number of haplotypes distinct to each respective species, that is without considering missing data (Ns) and ambiguous base-calls such as Y, W, N, etc. as differences in the assignment of each haplotype (for further details see the readme document at https://doi.org/10.5281/zenodo.4123683). Haplotype detection was performed on the total alignment with sequences generated in our study; for sequences with different lengths only the aligned length with sequence information for both sequences was considered for the distance calculation. Maximum and mean intraspecific distances and nearest-neighbor
genetic distances for all species were calculated using both PyCOIStats v1.3 and BOLD v4.

Sequences were aligned using the “BOLD aligner” option in BOLD v4, based on an Amino Acid Hidden Markov Model. The resulting alignment was checked by eye in Geneious R10\(^1\) and contained no insertions or deletions.

We used IQ-Tree v2.0.3 (Minh et al., 2020) to infer a maximum likelihood (ML) tree based on the alignment with distinct haplotypes and sequences > 500 bp. We used the IQ-Tree integrated ModelFinder with 286 DNA models, which selected GTR + F + R8 as the best-fit model according to the Bayesian information criterion. We then used the “monophylizer” perl script (Mutanen et al., 2016) on the IQ-Tree generated tree to assess species’ monophyly and test the concordance between COI-based and morphology-based species delineations.

Neighbor-joining (NJ) trees, used to visualize genetic distances within and between species and sequence clusters, were generated in MEGA X (Kumar et al., 2018) using uncorrected p-distances, and pairwise-deletion options. We used iTOL v.5.6.3 (Letunic and Bork, 2019) for the circular visualization of the NJ tree representing the entire DNA barcode library, after exclusion of all records of *Cam. ohridella* collected outside the native area of this invasive species (Valade et al., 2009).

Non-parametric Mann–Whitney *U* test was used to test for differences in haplotype diversity between alien and native species. Correlation between haplotype diversity and number of sequences per species was assessed with Pearson correlation (*p* < 0.05). These two statistical analyses were run in STATISTICA 8.0 (Stat Soft. Inc., United States).

**Identification of Candidate Species**

We used morphology, ecology (host plant use) and Barcode Index Numbers (BINs) to identify candidate species using an integrative approach (Figure 2A) (Padial et al., 2010). To identify species boundaries and assess the extent of overlooked diversity we used the candidate species approach of Vieites et al. (2009). We have used this integrative approach because it explicitly combines genetic, morphological and ecological data while systematics of leaf-mining moths traditionally considers host plant data in its workflow.
Each gracillariid record in the database was assigned to one of six categories shown in Figure 2A:

1. “BIN/species match”: there was a “one-to-one” match between a named species in the current taxonomy and the BIN assignment.
2. “Confirmed candidate species” (CCS): specimens assigned to a BIN (i.e., a divergent genetic cluster) that can be distinguished from currently described species by diagnostic morphological characters supporting the hypothesis that they represent an undescribed species; they are given a preliminary species name composed of the genus followed by the name of the host plant (if known) and that of the country (or one of the countries) where it is found, all written in one word and in between inverted commas.
3. “Unconfirmed candidate species” (UCS): specimens assigned to a BIN that represents a divergent genetic cluster for which morphological diagnostic characters remain tentative (either because available adult specimens were too few or not available), or for which differences in host plant use were found, indicating that they likely represent undescribed species. BINs in this category receive a preliminary species name as described above for CCS.
4. “Deep conspecific lineages” (DCL): specimens of a currently accepted species that are assigned to multiple BINs and for which detailed morphological and ecological comparative studies failed to reject the hypothesis that they belong to a single species.
5. “Unassessed species splits” (USS): specimens belonging to multiple BINs that have been assigned preliminarily to currently accepted species, but that could not be examined in detail within the framework of this study and could not be assigned to any of the three previous categories. We refer to those multiple BINs as “dark taxa.”
6. “Species complexes”: specimens belonging to morphologically distinct species that are not differentiated by DNA barcodes (shared BINs between species). These include unconfirmed synonymies (USY), where species that share the same BINs have been suggested by previous authors to be potential synonyms based on insufficiently distinct morphology.

Formal descriptions of new species are in preparation and will be published elsewhere.

FIGURE 2 (A) Workflow of integrative taxonomy (candidate species approach) used to identify species boundaries in European Gracillariidae records and to assign them to six species categories with main results. RESL, Refined Single Linkage Analysis implemented in BOLD; BIN, Barcode Index Number. (B) Number of BINs per species category (external circle in gray) and number of species within each category (inner circle): in green, European species lacking DNA barcodes; in beige, the number of named species in each of four species categories; in black, the potential “dark taxa” among 71 BINs representing the 26 USS; in red, the confirmed candidate species (CCS); in purple, the unconfirmed candidate species (UCS). Dotted lines delimit the three species categories comprising new (CCS) and potential cryptic diversity (UCS and USS/dark taxa) in European Gracillariidae. The question mark refers to the unknown number of cryptic species within the 71 BINs of USS category (we refer to those 71 BINs as dark taxa). In the 1:1 species/BIN and the CCS categories, the number of BINs (184) is slightly lower than species numbers (186), because two short DNA barcodes were not automatically assigned BIN numbers on BOLD; they nonetheless represent monophyletic genetic units and were considered within these categories. Percentages are given for each category with respect to a total number of species of 284 (including CCS and UCS), ignoring the unknown part represented by splits in unaddressed species (USS). (C) For the 263 named valid species recognized in 2020, this histogram represents the cumulative number of species described per decade since 1750. It also shows the 312 BINs and the increase from 263 described species (in green, 21 without DNA barcodes; in beige, 242 with DNA barcodes) to 284 species including 6 CCS (red) and 15 UCS (purple).
Laboratory Procedures and Analyses of ddRAD Data

We used ddRAD sequencing (Peterson et al., 2012) to test the hypothesis of cryptic diversity in five species with multiple BINs within a currently accepted species. We looked at whether there was any correspondence between BINs and RADseq clades in two species initially classified as USS [Phyllocnistis saligna (Zeller) and Phyllonorycter salicella (Zeller)] and three as DCL [Phyllonorycter roboris (Zeller), Ph. maestingella (Müller) and Ph. salicicolella (Sircom)]. Altogether 108 specimens with DNA barcodes were chosen for the ddRAD analyses, with a preference for samples that were recently collected at the time of DNA extraction. Extracts used for DNA barcoding which had been cryo-preserved at the Canadian Centre for DNA Barcoding (CCDB) were used for ddRAD library preparation.

Prior to the genomic library preparation, the quantity and quality of genomic DNA (gDNA) extracts were checked using PicoGreen Kit (Molecular Probes). To obtain a sufficient quantity of gDNA, whole genome amplification (WGA) was performed using REPLI-g Mini Kits (Qiagen) because of the low concentration of gDNA in the original extracts. As a result, 33 samples were omitted after WGA because they failed to meet minimum quality criteria (Supplementary Table S4). However, the reduction of specimens did not reduce the number of species tested. The ddRAD library was built following protocols described in Lee et al. (2018) with two exceptions: digestion was carried out using the restriction enzymes PsI and MspI (New England Biolabs) while the size distribution was controlled with a Bioanalyzer (Agilent Technology). The de-multiplexed fastq data are archived in the NCBI SRA: PRJNA562838.

We used ipyrad v.0.7.23 (Eaton and Overcast, 2016) to assemble loci de novo and create SNP datasets. All ipyrad defaults were used, with the following exceptions: the minimum depth at which majority rule base calls are made was set to 3, the cluster threshold was set to 0.80 and 0.90, and the minimum number of samples per locus was set to 2 and 3.

We inferred evolutionary relationships using ML methods implemented in RAxML v.8.2.0 (Stamatakis, 2014) with bootstrap support estimated by 1,000 replicates of rapid-bootstrap under the GTR + CAT model. We visualized the resulting phylogeny and illustrated bootstrap support using FigTree v.1.4.2 (Rambaut, 2015).

Detecting Wolbachia Infections

To test the hypothesis that Wolbachia was responsible for deep COI splits we looked at whether a particular species with deep splits showed a pattern of infected and uninfected BINs and or ddRADseq clades. Therefore, the same 108 specimens that were used for ddRAD sequencing were also screened for the presence of Wolbachia, using two markers, FstZ and Wsp. Primer sequences and PCR protocols were used according to Braig et al. (1998) and Zha et al. (2014); detailed information can be found in Supplementary Table S5.

Biogeography and Host Plant Data

We summarized the current known distribution of European gracillariid species based on the literature and our own records, assigning them to nine biogeographical categories: (1) Palearctic, (2) Euro-Siberian, (3) Euro-Central Asian, (4) European (5) Mediterranean, (6) Macaronesian, (7) Alpine, (8) Boreal, and (9) Holarctic (Supplementary Table S3). In addition, seven alien invasive species that have established recently in Europe have been included, i.e., three Nearctic and four Asian species (Supplementary Figures S1, S2).

We used BOLD to document and analyze the distribution range and host plant use of each of the multiple BINs within DCLs and their nearest neighbors (NNs). BINs were considered sympatric when both co-occur on the same site (i.e., also syntopically).

We compiled host plant data containing 2,368 host plant records belonging to 26 plant families, 81 plant genera and 206 plant species (Supplementary Table S3). In total, 1,594 host plant records are based on reared adults, 748 are based on barcoded immature stages (larvae or pupae) dissected from their leaf mines, and 26 are based on adults found resting or swept on their host plants (Supplementary Table S3). Supplementary Figure S3 shows host plant use for the whole European Gracillariidae fauna, including species not barcoded (Supplementary Table S6). We characterized the host plant range of each species as monophagous, oligophagous, or polyphagous (Hering, 1951) (Supplementary Table S3).

We standardized host plant names in the BOLD field “Associate Taxa,” largely following the 2020 Catalog of Life (Hassler, 2020), and likewise in Supplementary Table S3. We made the following exceptions:

1. We split the former Cytisus balansae (Boiss.) Ball (Fabaceae) into Cyt. balansae s. str., occurring in Spain (Granada and Almeria) and North Africa (European population previously known as Cyt. balansae subsp. nevadensis or Cyt. galianoi), and Cyt. oromediterraneus Rivas Mart et al., in other parts of Europe, following Auvray and Malecot (2013) (previously known as Cyt. purgans auct., Cyt. balansae var. europaicus).
2. The name Salix euxina Belyaeva is now used for what was previously known as Salix fragilis L., while the name S. fragilis is used for the much commoner hybrid of S. euxina with S. alba L. (Belyaeva, 2009). As separation of the full species and the hybrid is difficult and not very relevant for leaf miners, we retained the names (S. fragilis or S. euxina) as given by the original collector. Likewise, we kept the names S. cinerea L. and S. atrocineria Brot. as recorded by the original collector, as not all floras separate them and the differences are subtle. Most likely, the majority of records from the British Isles and western parts of Iberia and France in fact represent S. atrocineria.

In general, we accepted the host plant identification given by the original collector, where possible checking back with the collector in cases of doubt. Material of Ph. nevadensis Walsingham in the Deschka collection was labeled as being reared from Teline linifolia L. (now Genista linifolia L.). We consider this as a misidentification for Adenocarpus
decorticans Boiss., based on the locality, habitat and known host plant preference.

For host plant data from the literature we queried the Global Taxonomic Database of Gracillariidae (De Prins and De Prins, 2020), but while this could be comprehensive, it also includes old records that have no recent confirmations. In cases where such old records could lead to incorrect estimation of host plant range, we excluded them from Supplementary Table S3 (column AM) and Supplementary Table S6. An example is Sabulopteryx limosella (Duponchel), a known leaf miner of Turcium (Lamiaceae) for which old records of two unrelated hosts exist: Jurinea sp. (Asteraceae) and Genista sp. (Fabaceae). Rennwald in Lepiforum e.V. (2008–2020) critically considered these records and concluded that these must have been based on misidentifications.

Among the immature stages dissected from their leaf-mines, some interesting cases of xenophagy were noted, such as larvae of Cam. ohiidella found feeding on Corylus avellana (L.). This species has also been found feeding on Ostrya carpinifolia Scop. and Fagus sylvatica L. (Ellis, 2020). These cases of xenophagy have not been included in the global analysis of gracillarid/host plant interactions since they represent either oviposition errors or unusual host plants and do not reflect the “normal” host plant use of the leaf miner. This highlights the importance of distinguishing which host plant records are based on reared adults, as they better reflect the “true” host plant use of the leaf miner, versus those that are based on dissected immature stages which might contain oviposition errors.

RESULTS

DNA Barcoding Performance

We generated a total of 6,791 COI sequences representing 242 of the 263 species (92%) of Gracillariidae known to occur in Europe (Supplementary Table S2 and Supplementary Figure S4) with an average of 25.8 specimens per species [minimum one specimen (in 12 (5%) of the 242 named species), maximum 1,932 specimens] (Supplementary Table S3). We were unable to obtain sequences for 21 species (Supplementary Table S6). A total of 6,524 sequences (96.1%) were ≥500 bp long.

The monophylizer analysis identified 21 out of 242 named species (8.7%) species as non-monophyletic. The remaining 221 species (91.3%) are monophyletic and can therefore be identified reliably using barcodes (Supplementary Table S7). If we exclude 41 named species that are singletons (only one distinct haplotype per species) and by definition monophyletic, then the non-monophyly percentage is 10.5%.

Overall, 6,587 sequences were assigned by BOLD to one of the 312 different BINs (results on the 4th of November 2020) (Supplementary Table S3). Out of the 242 named species barcoded, 184 (76%) had their own unique BIN (one to one match category, Supplementary Table S3; Figure 2B). Only eight BINs out of the 312 (2.6%) were shared (seven BINs shared between species pairs and one BIN shared in a triad) among 17 species (Supplementary Table S3 column W). Definite species assignment using BINs is not possible in these 17 cases.

The dataset contains 1,491 distinct haplotypes with a maximum of 56 haplotypes found in 138 individuals of Ph. salicetella (Figure 3A) and an average of 5.7 haplotypes per species (Supplementary Table S3). Haplotype diversity was significantly correlated to sampling effort expressed in number of individuals sequenced per species (Pearson correlation: \( r = 0.70, N = 262, p < 0.05; \) Figure 3A).

We obtained barcode data for seven out of eight species alien to Europe (Supplementary Tables S3, S6), five European species that became invasive and expanded their range across Europe, and six European species alien in other continents (Supplementary Table S3). Stomphastis confluia (Gerasimov) has not yet been barcoded; it originates from north eastern tropical Africa and has been introduced to the Mediterranean with its host plant (Supplementary Table S6). The seven-alien species to Europe show a similar average number of haplotypes (5.7 ± 2.6) compared to the native European fauna (5.6 ± 0.4) (Mann–Whitney U test: \( Z = 0.3, p = 0.75, N_{alien} = 7 \) vs \( N_{native} = 256 \)). The alien species Ph. issikii shows one of the highest numbers of haplotypes (20) in our analysis (Figure 3A).

Removing non-distinct haplotypes prior to pairwise distance calculations and ignoring ambiguous bases in the comparisons with PyCOIStats scripts results in slightly larger differences (maximum intraspecific distance was 1.5%) (Supplementary Table S8) overall in comparison to a more conventionally used software, such as BOLD (maximum intraspecific distance was 1.4%) (Supplementary Table S3). In total, 46 species show maximum intraspecific distances higher than 2% and 12 species higher than 5% (Supplementary Table S3).

The pairwise distance statistics within and between species and BINs are shown in Figure 3B to allow for easy comparison with other DNA barcoding literature where such analyses are used to explore the presence or absence of a “barcoding gap.” DNA barcoding is deemed reliable if the variation within species (i.e., \( D_{max} \)) is smaller than the variation between species (i.e., \( D_{min\_NN} \)). For our dataset, both at the species and BIN level, there is some overlap between the two measures, indicating that a pairwise distance threshold to separate species cannot be set at a strict value. The maximum variation within species is greater when they are defined using classical taxonomy (max. 7.8%) compared with BINs (max. 3.4%), though this is to be expected as the BIN algorithm employs genetic distance data. Application of the BIN algorithm also leads to greater splitting and to smaller distances between sister species: on average 4.1% (BIN) versus 5.6% (species).

Species Discovery and Cryptic Diversity

Our integrative workflow combining morphology, host plant use data and DNA barcodes (Figure 2A) identified 21 new candidate species distinct from the 242 named species sampled (Figures 2B,C and Supplementary Table S3). Indeed, we discovered six confirmed candidate species (CCS) and 15 unconfirmed candidate species (UCS) that show a Mediterranean distribution (Figure 4), with eleven of these candidate species occurring on islands, four of them feeding on Alnus spp. (Supplementary Table S3). If we add these 21 new candidate species (6 CCS and 15 UCS) to the known European fauna of
gracillariids the total number of species raises from 263 to 284 species, an increase of 7.4% in species richness (Figures 2B,C).

In addition, 13 species were classified as deep conspecific lineages (DCLs), comprising a total of 27 BINs with no intraspecific morphological differences found, and no known ecological differentiation (Figure 2B and Supplementary Table S3). Twenty-six species were classified as "unassessed species splits" (Figures 2A,B); they comprise 71 BINs. More work is needed to assess the status of those 71 BINs that are considered here as "dark taxa".

The Mediterranean is the region with the highest species richness of gracillariids (mainly Lithocolletinae) in Europe (Figure 4 and Supplementary Figure S1) feeding mostly on Fabaceae and Fagaceae (Supplementary Figure S2 and Supplementary Table S3). The 15 Macaronesian endemic species feed on a wide variety of host plant families (Supplementary Figure S2) and their host plant range reflects the high local endemism of the flora (Hobohm, 2000) specializing on native endemic plants of high conservation value, e.g., Caloptilia laurifoliae (Hering) restricted to the native tertiary relic of Laurisilva forests (Supplementary Table S3).

Of the 39 species (13 DCL + 26 US) that show multiple BINs, 24 species (61.5%) have BINs with non-overlapping allopatric distribution ranges, whereas 11 species (28.2%) have BINs with overlapping sympatric distribution and four species (10.3%) have both allopatric and sympatric BINs (Supplementary Table S9).

Out of the 25 species with multiple BINs for which we have host plant data, 17 species (68%) have BINs that share the host plant, five species (20%) have BINs with non-overlapping host-plant ranges, and three species (12%) show a mix of BINs on separate and shared host plants (Supplementary Table S10). We identified nine species with different BINs co-occurring sympatrically and feeding on the same host plant (Supplementary Table S11).

The ddRAD sequencing of five species with multiple BINs produced a minimum of 663 loci (124,949 bp) and a maximum of 2,457 loci (465,525 bp) (Supplementary Table S4). Only one species, Phyllocnistis saligna, shows correspondence between the clades obtained via ddRAD ML trees and the BINs suggesting the presence of a cryptic species in the Czech Republic (Figure 5). However, sampling for ddRAD was limited and it is unclear if there is any ecological difference that correlates with the BINs. Moreover, BIN AAL5482, with European samples identified as Ph. asiatica solely based on association with this BIN, may prove to be conspecific with Ph. saligna when samples from across the Palearctic are combined in a single analysis.

Fourteen of the 108 (13%) samples analyzed were positive for infection by Wolbachia and all 14 infected samples were detected by the FtsZ gene (Supplementary Table S4). The pattern of infection is not associated with neither BIN clustering nor RADseq clades (Figure 5).

**DISCUSSION**

**DNA Barcoding as a Reliable Tool for Species Identification**

Over 73% of Lepidoptera species known to occur in Europe have been barcoded so far. Germany is leading the way on DNA barcoding progress in Europe with over 24,000 animal and plant species barcoded and an almost completed reference library for Lepidoptera, covering more than 3,300 species (91%) (Geiger et al., 2016; Hausmann pers. comm.). Identification reliability in the DNA barcode library of European
FIGURE 4 | DNA barcode variation in European Gracillariidae and geographical distribution of new candidate species. Neighbor-joining tree based on uncorrected p-distances for 5,148 records of the DNA barcode library of European Gracillariidae (1,643 records of Cameraria ohridella were removed to avoid topology distortion). The map shows all 6,791 georeferenced DNA barcode records for 242 named species and 21 new candidate species of European gracillariids. Colored circles represent individual records (map) or clusters (tree): in red six Confirmed Candidate Species (CCS); in purple 15 Unconfirmed Candidate Species (UCS); in orange 80 out of the 83 Mediterranean species sampled (three species, C. ohridella, Phyllonorycter platani, and P. leucographella were not colored on the map because they have undergone massive range expansion over most of Europe). Inner circle indicates the seven subfamilies of Gracillariidae found in Europe (Or: Ornixolinae; Oe: Oecophyllatinae; A: Acrocercopinae).

Lepidoptera is high, with levels of monophyly over 90% (Dincă et al., 2015; Mutanen et al., 2016). Our dataset shows similar levels of monophyly with only 8.7% of the 242 species sampled being non-monophyletic (10.5% if excluding species represented by singletons). Some cases of non-monophyly may still be attributable to operational error such as inaccurate reference taxonomy (unconfirmed synonyms, misidentifications) but others could be due to introgression/incomplete lineage sorting. For instance, two species pairs [Cal. laurifoliae (M. Hering)/Cal. nobilella (Klimesch) and Ph. adenocarpi (Staudinger)/Ph. parvifoliella (Ragonot)] have been hypothesized as synonyms. Another case revealed here is Ph. brunnea Deschka and Ph. acaciella (Duponchel). We obtained three short sequences of paratypes of Ph. brunnea which match the barcode of Ph. acaciella. Deschka (1975) described Ph. brunnea from moths reared from leaf mines on Ulmus that he collected on Rhodes (Greece). He noted that the genitalia are difficult to separate from those of the widespread Ph. acaciella that also
feeds on Ulmus, but external characters clearly differentiate the two. The mainland Ph. acaciella we sequenced do indeed appear much darker, with many dark scales obscuring the pattern (RMNH.INS.554149 and RMNH.INS.554150), compared to the barcoded paratypes of Ph. brunnea (CLV2670, CLV2671, and CLV2672). However, pale specimens of Ph. acaciella are sometimes found in mainland populations. Since the incomplete barcodes of the Ph. brunnea paratypes are inseparable from those of Ph. acaciella and we do not see real genitalia differences, we synonymize these two species, with Phyllonorycter brunnea Deschka (1975) syn. nov. becoming a junior synonym of Phyllonorycter acaciella (Duponchel, 1843).

Another case of non-monophyly is the "Ph. nicellii (Stainton)/Ph. chrysella (Constant)/Ph. froelichiella (Zeller)" complex, which could represent another potential case of mitochondrial introgression. Some specimens of Ph. froelichiella are placed within the Ph. nicellii BIN. Both species are clearly morphologically differentiated and feed on different host plant genera, so this could be another case of introgression. Both Ph. chrysella and Ph. froelichiella share the host plant Alnus glutinosa (Supplementary Table S3) and the three barcodes of Ph. chrysella are nested within Ph. nicellii and Ph. froelichiella. Another interesting case of non-monophyly and BIN-sharing is the Populus-feeding species complex of Phyllocnistis. This comprises the North American species P. populiella feeding mainly on Populus tremuloides Michx. and causing huge outbreaks in the Yukon and Alaska, and two European species P. labyrinthella (Bjerkander) developing mainly on Populus tremula L., but also frequently on Pop. alba L. and the hybrid Pop. × canescens (Aiton) Sm. (Ellis, 2020) and P. xenia M. Hering found mainly on Pop. alba (Aarvik et al., 2017). Although all three species share one single BIN, they show a low genetic divergence but consistent division between all three species. The nearest neighbor is a North American species feeding mainly on Populus balsamifera (Supplementary Table S3). We suspect that this complex represents a case of relatively
recent evolutionary divergence and we therefore treat them as separate species pending further study, particularly using nuclear markers and morphology.

Interestingly, DNA barcodes distinguish clearly between species that had previously been synonymized [for instance, Par. alpicola (Wocke)/Par. scoticella (Stainton); Ph. cerasicolella (Herrich-Schäffer)/Ph. spinicolella (Zeller); Triberta cistifoliella (Groschke)/T. helianthemella (Herrich-Schäffer)]. The validity of all these species is considered confirmed here.

Gracillariidae now joins the megadiverse Geometridae and Gelechiidae as one of the few large (2,000 + species globally) Lepidoptera families with a high coverage of barcoded species for Europe (Hausmann et al., 2013; Huemer et al., 2020). However, a total of 21 species remain to be barcoded to complete the DNA barcode reference library of known European gracillariids (Supplementary Table S6). Six out of the 21 species without barcodes occur in islands; nine do not have any existing host plant record (Supplementary Tables S6, S12). It will therefore likely be a tedious task to obtain fresh specimens of these missing species and protocols for barcoding old material might represent a preferable solution to further complete the barcode reference library (Prosser et al., 2016) from collection specimens.

During the DNA barcoding of both the 24 public and 36 privately owned reference collections we were able to detect numerous cases of misidentified specimens. Indeed, DNA barcodes helped to correct and curate the reference collections in particular for those adult moths that had been identified only on the basis of external morphology, and were collected at light without any host plant data associated.

Species Discovery and Cryptic Diversity

DNA barcoding has proved to be an efficient tool for accelerating species discovery of small moths both in tropical (Lees et al., 2014; Lopez-Vaamonde et al., 2019) and temperate areas (Mutanen et al., 2013; Huemer et al., 2020). By combining DNA barcodes, morphological and host plant data, our study shows the diversity of Gracillariidae in Europe to have also been underestimated, and that there may be as many as 284 species in Europe (considering both CCS and UCS revealed here), possibly many more if we take into account the 45 unaddressed DNA barcode splits within USS, called here dark taxa (Figure 2B).

We found 21 lineages that represent candidate new species (6 CCS and 15 UCS). These newly revealed candidate species occur in the mountainous regions of Mediterranean islands, the South-Eastern Alps and the Balkans. These candidate species are likely to be rare, and knowing their true geographical distribution is crucial for their effective conservation. Our results highlight the Mediterranean as the most important biogeographical region within Europe in terms of gracillariid species richness, number of endemic species and new undescribed taxa.

Our analyses have also revealed 39 (16%) species with multiple BINs (i.e., deep COI splits; DCL and USS) and some of those might represent cryptic species. A barcoding study on European Gelechiidae micromoths found a similar level (15%) of species with multiple BINs (Huemer et al., 2020). However, these results will vary depending on sampling efforts, especially at intraspecific level. Indeed, most large-scale DNA barcode studies favored maximizing the number of different taxa sampled and neglected intraspecific sampling with typically fewer than ten individuals per species (Phillips et al., 2019). Our sampling effort with an average of 20.5 sequences per BIN significantly exceeds the average barcoding study. However, in our dataset, 63.5% (198/312) of the BINs were composed of less than 10 records, indicating a need to improve sampling efforts. One way of increasing intraspecific sampling is by combining national and regional DNA barcoding libraries. For instance, Finland had its entire gracillariid fauna DNA barcoded. Other countries show good progress such as Austria with 83.9% DNA barcoded species (119 barcoded out of 137 recorded species). On the contrary, Ireland and Estonia, for example, lack gracillariid DNA barcodes completely, thus efforts should be focused on such areas to extend coverage by sampling, if not of species.

Because the investigation of each individual case of DNA barcode divergence within species requires much time and material, our study leaves a total of 71 BINs within 26 named species as “unassessed species splits” (USS). Although they currently are assigned a valid Linnean species name in our reference library, we refer to those BINs as “dark taxa” (Page, 2016), until their status as potential cryptic species is fully examined. It was well beyond the scope of this study to examine all those multiple BINs, but we used nuclear data to explore the deep COI splits of two USS and three DCLs. Results obtained from the analysis of the ddRADseq data supports the presence of cryptic species in only one (Phyllocnistis saligna) of the five species studied. Indeed, six P. saligna specimens of BOLD:AAQ1589 and three specimens of BOLD:ABW1114 form different RADseq clusters (Figure 5). Additional specimens must be both barcoded and RAD sequenced to confirm this result. Therefore, we keep this species in the category “unassessed species splits” (Supplementary Table S3). While this result might appear to undermine the idea that intraspecific DNA barcodes splits are frequently indicative of cryptic diversity, it is likely biased because of the often-large proportion of missing data that characterize datasets generated with this method (Dincă et al., 2019), particularly when studied taxa are not closely related (Lee et al., 2018). Further genome-wide DNA analyses could be carried out using universal single-copy orthologs (USCOs) (Eberle et al., 2020) and ultra-conserved elements (UCEs) (Gueuning et al., 2020) coupled with breeding experiments to understand isolating mechanisms (Ohshima, 2005, 2008; Ohshima and Yoshizawa, 2010).

Another mechanism that could lead to the split of a species into deep mitochondrial lineages is via Wolbachia mediated spread of a new mitochondrial haplotype after an introgressive event. Indeed, deep mitochondrial splits have been found to be associated with specific Wolbachia strains in butterflies (Kodandaramiah et al., 2013; Ritter et al., 2013; but see Bereczki et al., 2015). West et al. (1998) found that 38% (8/21, all but one Phyllonorycter spp.) of gracillariids analyzed were infected with Wolbachia, while Gutierrez et al. (2015) found a similar percentage, 30% (35/119), in a larger sample. In our study only 13% of specimens analyzed (14/108) were found to be infected. This low level of prevalence must be taken with caution because we used only two molecular markers to detect Wolbachia and we screened a relatively small number of individuals per species. In addition, we did not find examples of where Wolbachia infection...
was associated with one particular BIN or RADseq clade within a morphospecies. This suggests that the endosymbiont does not play a major role in the generation of deep splits but further sampling is needed to confirm this conclusion. The possibility that other mitochondrial drive elements may be involved should also be considered.

Only one species, *Phyllocnistis saligna*, showed correspondence between nuclear and mitochondrial splits (Figure 5A). The two Czech specimens that show both nuclear and mitochondrial divergence are morphologically undistinguishable to the other European material examined from Austria, Germany and Finland. This might suggest the potential presence of a cryptic species in the Czech Republic but further specimens must be examined to confirm this hypothesis.

In our study we also looked at whether different BINs would correlate with geography and/or host plant shifts. We found that over 24 of the 39 species (61.5%) that show intraspecific splits have allopatric BINs (Supplementary Table S9). This pattern of checkered distributions has also been shown among cryptic taxa in other Lepidoptera like butterflies (Vodá et al., 2015). However, the number of individuals for many BINs is relatively low and it is likely that the number of BINs with sympatric distribution will increase as sampling effort increases.

Our analyses show that 68% (17/25) of the species with multiple BINs for which we have host plant data have BINs that share the host plant. This would suggest that host plant shifts are not important drivers of divergence in gracillariids (Imada et al., 2011). On the other hand, some studies have found host plant shifts between distantly related plants as the major driver of speciation in gracillariids (Ohshima, 2008).

We also identified nine species with multiple BINs co-occurring sympatrically and feeding on the same host plant (Supplementary Table S11). These species are good candidates to study *Wolbachia*-mediated genetic sweeps since neither geography nor host plant shifts seem to explain their COI splits.

To measure the relative impact of geography and host plant shifts on the diversification of gracillariids, we would need a robust phylogeny on which to reconstruct the major ancestral host switches and distributions (Winkler et al., 2018). The latest molecular phylogenetic study revised the classification of Gracillariidae into eight subfamilies, but relationships between them remain largely uncertain (Kawahara et al., 2017).

**CONCLUSION**

We have assembled a large barcode dataset covering nearly all of the European gracillariid fauna, allowing 91.3% of the known and named species to be identified accurately. We discovered new candidate species in Mediterranean mountains and islands. We also revealed a large number of DNA barcode splits (71 BINs in 26 US), that have not been investigated through an integrative approach yet and may comprise many potentially cryptic lineages. Further assessment of those cases represents a major task that will require additional sampling, comparative morphology studies, breeding experiments or genome wide analyses.

The DNA barcode library made available by this study will enable identification of any life-stage, help discover and document the true diversity of European gracillariids, and contribute to the conservation assessment of these fascinating insects.

**DATA AVAILABILITY STATEMENT**

The resulting sequences, along with the voucher data, images, and trace files from Sanger sequencing, are deposited in the Barcode of Life Database (BOLD), http://www.boldsystems.org (Ratnasingham and Hebert, 2007) and the sequences subsequently deposited in GenBank. All data are available from BOLD in the dataset DS-EUGRACI2 (dx.doi.org/10.5883/DS-EUGRACI2). Github project DOI: https://doi.org/10.5281/zenodo.4123683 'PyCOIstats' contains the scripts used for analyzing the COI sequences.

**AUTHOR CONTRIBUTIONS**

CL-V and RR designed the study, carried out most of the analyses, and wrote the first draft. EN double checked and validated the host plant data and nomenclature. AG, NK, RR, and SG made the figures. NK analyzed the biogeographical and host plant data. CD ran both the PyCOIstats and monophyly tests. KL carried out the molecular lab work of ddRAD, performed the bioinformatics, data analysis, and drafted the ddRAD section. CL-V, AC, CD, HG, SG, PH, NK, J-FL, AL, ZL, MM, EV, AS, PT, and CW contributed to the material collection and DNA barcoding. CL-V led the writing of the manuscript. RR, NK, EV, AG, HG, CD, and DL contributed substantially to revisions. All authors contributed to the article and approved the submitted version.

**FUNDING**

This work was funded by the Studium (France) to both NK and DL to work as research fellows at INRAE Orléans; the Russian Foundation for Basic Research (No. 19-04-01029-a) and the basic project of Sukachev Institute of Forest SB RAS (project No. 0287-2021-0011) to NK; the Région Centre (France) project ENDOFEED (201000047141) to CL-V; the Ministry of Agriculture (France) project PASSIFOR to CL-V and RR; the Academy of Finland (grant number: 277984) to MM and the Kvantum Institute to KL; Kone foundation, Finnish Cultural foundation for financing the FinBOL (Finnish Barcode of Life) 2011-2015, and the Academy of Finland for financing FinBOL through the FinBIF national infrastructure project since 2014; Agriculture & Agri-Food Canada, Science and Technology Branch (projects Nos. J-000977 and J-002279) to J-FL; Promotion of Educational Policies, University and Research Department of the Autonomous Province of Bolzano – South Tyrol and Austrian Federal Ministry of Science, Research and Economics in the framework of ABOL (Austrian Barcode of Life) to PH. CD’s salary and all DNA sequencing at Naturalis Biodiversity Center was funded through the Dutch
Economic Restructuring Fund (FES). AS was supported by the Bavarian Ministry of Science and Art (Germany) project BFB and the Federal Ministry of Education and Research (Germany) project GBOL. Funding for DNA barcoding was also partly provided by the Government of Canada through Genome Canada, the Ontario Genomics Institute in support of the International Barcode of Life project, and by the National Science and Engineering Research Council of Canada.

ACKNOWLEDGMENTS

We are grateful to Ian Barton, Bengt Åke Bengtsson, Sylvfied Charlat, Helmut Deutsch, Jurate and Willy De Prins, Gertfried Deschka (deceased), Willem Ellis, Gerald Fuchs, Theo Grünwald, Thomas Guggemoos, Alfred Haslberger, Axel Hausmann, Richard Heindel, Povilas Ivinskis, John Langmaid, Peter Lichtmannnecker, Wouter Moerland, Enrique Murria, Elisenda Olivella, Annette von Scholten-PfP, Ian Sims, Franz Theimer, Andreas Werno, Werner Wolf, and Steve Wullaert for sending valuable samples and/or allowing access to their collections. Thanks to Leif Aarvik, Kai Berggren, Jeremy deWaard, Per Falck, Arild Johnsen, Jari Junninainen, Mark Sterling, and Martin Sigut for allowing access to unpublished barcodes. Thanks to both Ole Karsholt and Svetlana Baryshnikova for discussion on Microapteryx bistrigella and Caloptilia alpherakiella respectively. We would like to thank Scott Miller for allowing us to use 33 sequences obtained as part of a collaboration between the Smithsonian Institution and the University of Guelph to barcode historical Lepidoptera specimens. We are grateful to Laura Tormälä for her efficient work in the laboratory at University of Oulu. The authors also wish to acknowledge CSC–IT Centre for Science, Finland and the technical support and advanced computing resources from University of Hawaii Information Technology Services – Cyberinfrastructure for providing computational resources. Thanks to two referees for their insightful comments.

This article is dedicated to the memory of the late Gertfried Deschka for his major contribution to the knowledge of European Gracillariidae and warm hospitality during our (CL-V, DL, and SG) visits to Steyr to work on his extensive collection and his help during this project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2021.626752/full#supplementary-material

REFERENCES

Aarvik, L., Bengtsson, B. Å, Elven, H., Ivinskis, P., Jüritve, U., Karsholt, O., et al. (2017). Nordic-baltic checklist of lepidoptera. Nor. J. Entomol. Suppl. 3, 1–236.

Avray, G., and Malecot, V. (2013). A revision of Cytisus sections Alburnoides, Spartopis and Verzinnun (Genisteae, Fabaceae). Edinburgh J. Bot. 70, 61–120. doi: 10.1017/S0960428612000352

Belyaeva, I. (2009). Nomenclature of Salix fragilis L. and a new species, S. euxina (Salicaceae). Taxon 58, 1344–1348. doi: 10.1002/tax.58402

Bengtsson, B. Å, and Johansson, R. (2011). Nationalnyckeln till Sveriges flora och fauna. Fjärilar: Bronsmalar-rullvingemalar (Lepidoptera: Roeslerstammiidae, Lyonetidae. ArtDatabanken, SLU, Uppsala.

Berezki, J., Racz, R., Varga, Z., and Toth, J. P. (2015). Controversial patterns of Wolbachia infestation in the social parasitic Maculinea butterflies (Lepidoptera: Lycaenidae). Organ. Div. Evol. 15, 591–607. doi: 10.1007/s13127-015-0217-7
Deschka, G. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.

Dinc˘a, V., Lee, K. M., Vila, R., and Mutanen, M. (2019). The conundrum
of species delimitation: UCEs resolve long-standing systematic enigma in Central
European bees. Genome
59, 661–670. doi: 10.1139/gen-2015-2185

Dinc˘a, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L.,
Garcia-Barros, E., et al. (2015). DNA barcode reference library for Iberian
butterflies enables a continental-scale preview of potential cryptic diversity. Sci.
Rep. 5:12395. doi: 10.1038/srep12395

Dinc˘a, V., Wiklund, C., Lukhtanov, V. A., Kodandaramaiah, U., Noren, K.,
Dapparto, L., et al. (2013). Reproductive isolation and patterns of genetic
differentiation in a butterfly species complex. J. Evol. Biol. 26, 2095–
2106. doi: 10.1111/jeb.12211

Doorenweerd, C., van Haren, M. M., Schermer, M., Pieterse, S., and van
Nieuken, E. J. (2014). A Linaeaus NG (TM) interactive key to the
Lithocolettinae of North-West Europe aimed at accelerating the accumulation
of reliable biodiversity data (Lepidoptera, Gracillariidae). Zookeys 422, 87–101.
doi: 10.3897/zookeys.422.7446

Doorenweerd, C., van Nieuken, E. J., and Hoare, R. J. B. (2017). Phylogeography,
classification and divergence times of pygmy leaf-mining moths (Lepidoptera:
Nepticulidae): the earliest lepidopteran radiation on Angiosperms? Syst.
Entomol. 42, 267–287. doi: 10.1111/syen.12212

Draisma, G. K. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.

Draisma, G. K. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.

Draisma, G. K. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.

Draisma, G. K. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.

Draisma, G. K. (1975). Blattminierende Lepidopteren aus dem Nahen und Mittleren
Osten. I. Teil. Zeitschrift der Arbeitsgemeinschaft Österreichischer Entomologen
26, 41–46.
Kirichenko, N., Triberti, P., Kobayashi, S., Hirowatari, T., Doorenweerd, C., Ohshima, I., et al. (2017). Systematics of Phyllonorycter leaf-mining moths (Lepidoptera, Gracillariidae) feeding on dogwood (Cornus spp.) in Northeast Asia, with the description of three new species. ZooKeys 736, 79–118. doi: 10.3897/zookeys.736.20739

Kirichenko, N., Triberti, P., Ohshima, I., Hanan, J., Byun, B.-K., Li, H., et al. (2017). From east to west across the Palearctic: phylogeography of the invasive lime leaf miner Phyllonorycter issikii (Lepidoptera: Gracillariidae) and discovery of a putative new cryptic species in East Asia. PLoS One 12:e0171104. doi: 10.1371/journal.pone.0171104

Kodandaramaiah, U., Simonsen, T. J., Bromilow, S., Wahlberg, N., and Sperling, F. (2013). Deceptive single-locus taxonomy and phyogeography: wolbachia-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. Ecol. Evol. 3, 5167–5176. doi: 10.1002/ece3.886

Kozlov, M. V., Mutanen, M., Lee, K. M., and Huemer, P. (2017). Cryptic diversity in the long-horn moth Nemophora degeerella (Lepidoptera: Adelidae) revealed by morphology, DNA barcodes and genome-wide dRAD-seq data. Syst. Entomol. 42, 329–346. doi: 10.1111/syen.12216

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X. Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. doi: 10.1093/molbev/msy096

Langmaid, J., and Corley, M. (2007). Phyllocnistis ramulicola sp. nov (Lepidoptera: Gracillariidae) in England and Portugal. Ent. Gazette 58, 227–237.

Laitütkä, Z., Laitütkä, A., and Lopez-Vaamonde, C. (2013). A revision of the Phyllonorycter ulicicolella species group with description of a new species (Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lee, K. M., Kivéla, S. M., Ivanov, V., Hausmann, A., Kaila, L., Wahlberg, N., et al. (2018). Information dropout patterns in restriction site associated DNA phylogenomics and a comparison with multilocus Sanger data in a species-rich moth genus. Syst. Biol. 67, 925–939. doi: 10.1093/sysbio/syy029

Lees, D. C., Kawahara, A. Y., Bouteleux, O., Ohshima, I., Kawakita, A., Rougerie, R., et al. (2014). DNA barcoding reveals a largely unknown fauna of Gracillariidae moth genus. Phyllonorycter issikii (Lepidoptera: Gracillariidae) and discovery of a putative new cryptic species in East Asia. PLoS One 12:e0171104. doi: 10.1371/journal.pone.0171104

Lepiforum e.V. (2008–2020). Lepiforum’s Checklist of European Lepidoptera, version 7. Lepiforum e.V.

Lees, D. C., Lack, H. W., Rougerie, R., Hernandez, A., Raus, T., Avtzis, N. D., Lie, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. doi: 10.1093/molbev/msy096

Lees, D. C., Lack, H. W., Rougerie, R., Hernandez, A., Raus, T., Avtzis, N. D., Lie, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. doi: 10.1093/molbev/msy096

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates. Bioinformatics 35, W256–W259.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.

Lepidoptera: Gracillariidae). SHILAP Rev. Lepidopt. 41, 251–265.
Vieites, D. R., Wollenberg, K. C., Andreone, F., Köhler, J., Glawe, F., and Vences, M. (2009). Vast underestimation of Madagascar’s biodiversity evidenced by an integrative amphibian inventory. *PNAS* 106, 8267–8272. doi: 10.1073/pnas.0810821106

Vodă, R., Dapporto, L., Dincă, V., and Vila, R. (2015). Why do cryptic species tend not to co-occur? A case study on two cryptic pairs of butterflies. *PLoS One* 10:e0117802. doi: 10.1371/journal.pone.0117802

West, S. A., Cook, J. M., Werren, J. H., and Godfray, H. C. J. (1998). *Wolbachia* in two insect host-parasitoid communities. *Mol. Ecol.* 7, 1457–1465. doi: 10.1046/j.1365-294x.1998.00467.x

Winkler, I., Scheffer, S. J., Lewis, M. L., Ottens, K. J., Rasmussen, A. P., Gomes-Costa, G. A., et al. (2018). Anatomy of a Neotropical insect radiation. *BMC Evol. Biol.* 18:9. doi: 10.1186/s12862-018-1146-9

Zha, X., Zhang, W., Zhou, C., Zhang, L., Xiang, Z., and Xia, Q. (2014). Detection and characterization of *Wolbachia* infection in silkworm. *Genet. Mol. Biol.* 37, 573–580. doi: 10.1590/S1415-475720140001400014

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Lopez-Vaamonde, Kirichenko, Cama, Doorenweerd, Godfray, Guiguet, Gomboc, Huemer, Landry, Laštůvka, Laštůvka, Lee, Lees, Mutanen, van Nieukerken, Segerer, Triberti, Wieser and Rougerie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.