diversity

MDPI

Article

An Incomplete European Barcode Library Has a Strong Impact on the Identification Success of Lepidoptera from Greece

Peter Huemer [1,*] and Marko Mutanen [2]

[1] Naturwissenschaftliche Sammlungen, Sammlungs- und Forschungszentrum, Tiroler Landesmuseen Betriebsges.m.b.H., 6060 Hall in Tirol, Austria
[2] Ecology and Genetics Research Unit, University of Oulu, 90014 Oulu, Finland; marko.mutanen@oulu.fi
* Correspondence: p.huemer@tiroler-landesmuseen.at; Tel.: +43-51259489-721

Abstract: Species identification by means of DNA barcodes depends essentially on the scope and quality of a relevant reference library. The first analysis of a large number (about 600 morphospecies) of southern European Lepidoptera (Greece: Peloponnese) shows both the advantages and disadvantages with regard to a reliable identification of Mediterranean species. We determined 946 DNA barcode sequences from 47 families, of which 929 sequences from 46 families were successfully assigned to a Barcode Index Number (BIN) in the global Barcode of Life Data Systems (BOLD) database. A species level identification for 485 BINs representing 477 Linnaean names was successful. These taxa include 34 new records for Greece. However, 128 BINs (c. 20% of the inventory) could not be attached to a Linnaean name from referenced sequences available in BOLD. Of these BINs, 99 are new and hence represent unique records for BOLD. Intra- and inter-BIN divergences are presented and discussed. An initial and preliminary in-depth analysis of randomly selected species indicates an incomplete DNA barcode library in terms of Linnean taxa, in addition to a considerable number of probably undescribed species. It is therefore strongly recommended that the already advanced European barcode library of Lepidoptera should be supplemented with not-yet-sequenced taxa from the Mediterranean.

Keywords: Barcode Index Number (BIN); DNA barcoding; faunistics; Mediterranean; new species; unidentified species

Citation: Huemer, P.; Mutanen, M. An Incomplete European Barcode Library Has a Strong Impact on the Identification Success of Lepidoptera from Greece. *Diversity* **2022**, *14*, 118. https://doi.org/10.3390/d14020118

Academic Editor: Stephan Koblmüller

Received: 29 December 2021
Accepted: 4 February 2022
Published: 7 February 2022

1. Introduction

During the last decade, considerable efforts have been made in several European countries to establish mostly national DNA barcode libraries of regional faunas, with only Finland so far having published a comprehensive reference library for the arthropod fauna [1]. Lepidoptera are furthermore almost completely covered in Norway, The Netherlands, Germany and Austria, although published results only deal with parts of the fauna [2–5], leaving numerous sequences for private use only. In addition to national initiatives, only a few studies have dealt with larger biogeographical regions, the project on the Lepidoptera of the Alps by the Tyrolean State Museums, Innsbruck, Austria being an exception [6]. Furthermore, selected larger families or superfamilies have been genetically studied on a continental scale by means of DNA barcodes, in particular the Gelechiidae [7,8], Gracillariidae [9], large parts of the Geometridae [10] and Papilionoidea [11]. With about 40 K barcodes, the largest dataset on Lepidoptera has been released by Mutanen et al. [12]. Apart from a few individually funded research projects, the vast majority of the sequences have been, and are being, incorporated into the global, publicly accessible Barcode of Life Data Systems (BOLD; [13]). About three-quarters of around 10,700 species of Lepidoptera known from Europe [14] are represented in BOLD by at least one reference sequence, while around 2600 species, typically very rare, do not yet have a reference DNA barcode. Due to the national initiatives mentioned above, DNA barcodes of almost all species in Northern and Central Europe can be assigned to Linnaean names or already known genetic clusters.

The extent to which these favorable conditions also affect the identification success of southern European samples was tested here on a larger scale for the first time using barcode data from more than 600 morphospecies from Greece.

2. Materials and Methods

A representative portion of Lepidoptera species from Greece was collected during two excursions to the Peloponnese peninsula, from 9 to 25 May 2019 and 9 to 29 September 2020. Collecting efforts covered all taxonomic groups except for butterflies, which have already been studied on a continental level [11]. Various survey methods, in particular light capture, were used to sample a representative species spectrum in the selected study areas on the Peloponnese peninsula. A priority goal was the sampling of about two individuals on average from the most possible number of species and thus from all lepidopteran families, including the so-called microlepidoptera. Of those taxa that are known to be difficult to distinguish from congeners, up to five specimens were collected. Species differentiations were provisionally carried out in the field according to morphological criteria. The material was pinned directly on site and immediately dried for further investigations. The final sample selection was later carried out in the laboratory according to external characteristics such as wing markings and color, head characteristics and occasionally also after preliminary examination of the genital morphology.

Finally, tissue samples (dried legs) of 1056 specimens of provisionally identified morphospecies were prepared according to the prescribed standards to obtain DNA barcode sequences of the mitochondrial COI gene (cytochrome c oxidase 1) [15]. Samples covered 47 families, namely, the Gelechiidae (n = 166), Geometridae (n = 136), Noctuidae (n = 123), Pyralidae (n = 101), Crambidae (n = 63), Erebidae (n = 57), Tortricidae (n = 43) and another 40 families (n = 367).

Material was processed at the Canadian Centre for DNA Barcoding (CCDB, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada) using the standard high-throughput protocol described in deWaard et al. [15]. Details including complete voucher data and images can be accessed in the public dataset "Lepidoptera Barcoding Greece" (https://dx.doi.org/10.5883/DS-LEPGREEC in BOLD, accessed on 28 December 2021). Sequences were then submitted to GenBank.

All sequences were assigned to the Barcode Index Numbers (BINs), algorithm-based operational taxonomic units that provide an accurate proxy for the true species [16]. BINs were automatically calculated for records in BOLD that are compliant with the DNA Barcode standard [17]. As BOLD presently does not have a functionality of calculating intra- and inter-BIN divergences, those values were provided for us by BOLD support. Follow-up species identification strictly followed available reference sequences in BOLD with a cross-check control of external morphology. In the case of BINs covering more than one taxon in BOLD (BIN-sharing, misidentifications, contaminations), identification was based on external morphology and, in critical cases, also on genital morphology. BINs attributed to a single Linnean name were accepted as correct although potential misidentifications cannot be fully ruled out.

Degrees of intra- and interspecific variation in DNA barcode fragments were calculated under the Kimura 2 parameter model of nucleotide substitution using analytical tools of BOLD systems v. 4.0. Finally, a Neighbor-Joining tree was constructed from these data in Newick format and a tree was drawn using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/, accessed on 28 December 2021).

3. Results

3.1. General Overview

Sequencing of 1056 tissue samples resulted in 946 DNA barcode sequences (c. 90% success rate). Full barcodes of 658 bp were recovered for 733 specimens, and for a further 167 specimens a sequence ranging between 600 and 657 bp was recovered. Sequences less

than 600 bp were obtained for only 46 specimens. Of all records with sequence data recovered, 834 sequences were considered to be barcode compliant following BOLD standards.

The vast majority of 929 sequences was assigned to a total of 614 different BINs in BOLD, leaving only 17 sequences without a BIN (see Figure S1). Of all BINs, 535 belong to only 15 families, the remaining 79 BINs representing 31 usually smaller families (Figure 1, Table S1). For one family (Adelidae) only a short sequence without a BIN was recovered. Average intra-BIN variability ranged from 0% to a maximum of 1.98% (mean 0.44%), and maximum intra-BIN distances ranged from 0% to 5.41% (mean 1.18%). Distances to nearest neighbor BINs ranged from a minimum of 1.04% to a maximum of 13.14% (mean 4.16%) (Table S1).

3.2. Unidentified BINs—Incomplete Barcode Library

A total of 128 BINs could not be attached to a Linnaean name from referenced sequences available in BOLD. Of these BINs, 99 are new and hence represent unique records for BOLD, whereas the remaining 29 BINs have been recorded from elsewhere (Table S1).

Judging from morphology, the unidentified BINs probably represent separate taxa at species level belong to 27 families. The Autostichidae with 17 and Gelechiidae with 14 unidentified BINs are the most important families in this respect, followed by Noctuidae, Pyralidae and Tineidae, each with nine, and Coleophoridae and Erebidae, each with seven unidentified BINs. Cosmopterigidae, Crambidae, Geometridae and Tortricidae have six unidentified BINs each, and the remaining 16 families have between one and four unidentified BINs (Figure 1).

An unknown portion of the so-far unidentified BINs is based on an incomplete barcode library of European Lepidoptera. The lack of reference sequences in BOLD thus prevents unambiguous identification. Time-consuming morphological investigations are therefore necessary in order to ultimately assign BINs to Linnaean names. Through such research, several taxa have already been identified in the context of the present study that did not have relevant reference sequences in BOLD, e.g., *Nukusa cinerella* (Rebel, 1891), *Aglossa signicostalis* Staudinger, 1870, and *Paidia cinerascens* (Herrich-Schäffer, 1847).

3.3. Potential Cryptic Diversity

The BINs that have not yet been assigned to species level certainly also include previously overlooked cryptic species, but the fraction of nameless taxa cannot be estimated as they require meticulous taxonomic scrutiny on a case-by-case basis. Examples of cryptic diversity can be found in the genera *Ypsolopha*, *Aristotelia* and *Coleophora*, where preliminary analysis of morphology indicates new taxa. However, comprehensive taxonomic investigations are required for further clarification.

3.4. BINs Attributed to Linnaean Names

BOLD identification analytics resulted in a species level identification for 485 BINs representing 477 Linnaean names (Table S1). Eight species with more than one sequenced specimen from Greece showed large genetic variation and were consequently clustered in two BINs. Such cases require in-depth analysis for possible cryptic diversity. Furthermore, two species, viz. *Cryphia ochsi* (Boursin, 1940) and *Trichoplusia circumscripta* (Freyer, 1831) shared their BINs with other taxa in BOLD, for which reason identifications were made based on external and genital morphology.

The Noctuidae with 87 identified species (18% of identified BINs) are followed by Geometridae (71 spp.), Gelechiidae (57 spp.), Pyralidae (46 spp.), Erebidae (36 spp.), Crambidae (32 spp.), Tortricidae (27 spp.) and another 35 families (Figure 1).

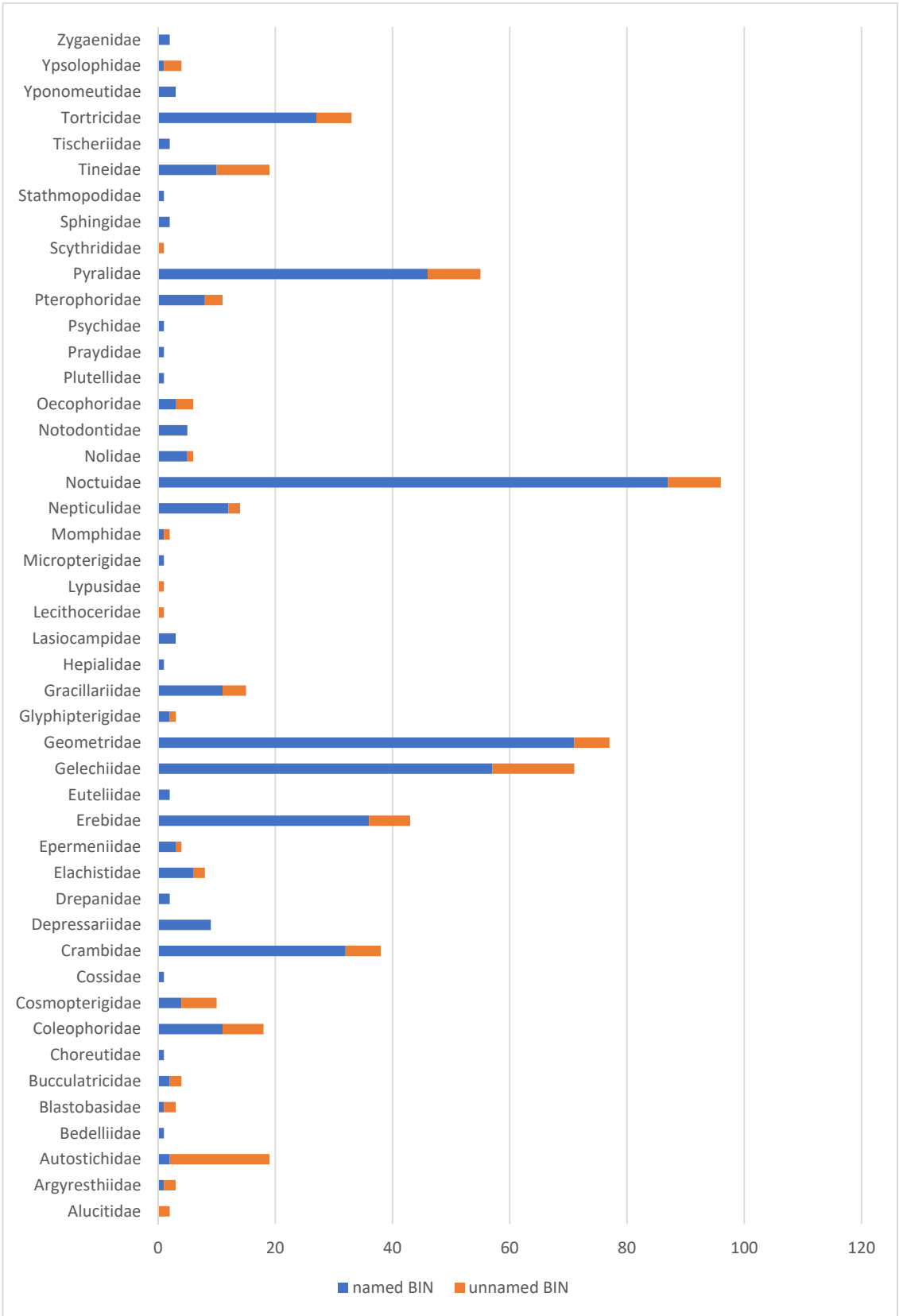

Figure 1. Named and unnamed BINs in a sample of Lepidoptera from Greece.

3.5. New Faunistic Records

Following refs. [18,19] 34 species attributed to a Linnaean taxon had not been recorded from Greece (Table 1). Further, three species, viz. *Helcystogramma lamprostoma* (Zeller, 1847), *Ornativalva heluanensis* (Debski, 1913) and *Caradrina levantina* Hacker, 2004 are new records for the Greek mainland. With the exception of *Eupithecia ultimaria* Boisduval, 1840 and *Bryophila felina* (Eversmann, 1852) all the remaining species belong to 11 families of so-called microlepidoptera, which are obviously under-represented in faunistic papers from Greece.

Table 1. New faunistic records for Greece.

Taxon	Family
Coleophora helgada (Anikin, 2005)	Coleophoridae
Coleophora gardesanella Toll, 1953	Coleophoridae
Eteobalea siciliae (Riedl, 1966)	Cosmopterigidae
Agriphila brioniellus (Zerny, 1914)	Crambidae
Friedlanderia cicatricella (Hübner, 1824)	Crambidae
Spoladea recurvalis (Fabricius, 1775)	Crambidae
Elachista argentella (Clerck, 1759)	Elachistidae
Elachista atricomella Stainton, 1849	Elachistidae
Elachista biatomella (Stainton, 1848)	Elachistidae
Elachista chrysodesmella Zeller, 1850	Elachistidae
Elachista obliquella Stainton, 1854	Elachistidae
Anarsia leberonella Réal, 1994	Gelechiidae
Aproaerema sangiella (Stainton, 1863)	Gelechiidae
Aristotelia subdecurtella (Stainton, 1859)	Gelechiidae
Ivanauskiella occitanica (Nel & Varenne, 2013)	Gelechiidae
Mesophleps oxycedrella (Millière, 1871)	Gelechiidae
Scrobipalpa superstes Povolný, 1977	Gelechiidae
Thiotricha subocellea (Stephens, 1834)	Gelechiidae
Eupithecia ultimaria Boisduval, 1840	Geometridae
Stigmella obliquella (Heinemann, 1862)	Nepticulidae
Bryophila felina (Eversmann, 1852)	Noctuidae
Batia inexpectella Jäckh, 1972	Oecophoridae
Acrobasis bithynella Zeller, 1848	Pyralidae
Acrobasis fallouella (Ragonot, 1871)	Pyralidae
Assara concicolella (Constant, 1884)	Pyralidae
Ceutolopha isidis (Zeller, 1867)	Pyralidae
Phycita torrenti Agenjo, 1962	Pyralidae
Tischeria decidua Wocke, 1876	Tischeriidae
Clepsis burgasiensis (Rebel, 1916)	Tortricidae
Cydia rymarczyki Varenne & Nel, 2013	Tortricidae
Lobesia bicinctana (Duponchel, 1844)	Tortricidae
Neocochylis dubitana (Hübner, 1799)	Tortricidae
Pammene herrichiana (Heinemann, 1854)	Tortricidae
Ypsolopha alpella (Denis & Schiffermüller, 1775)	Ypsolophidae

4. Discussion

Due to extensive recent achievements in the implementation of a DNA barcode reference library, the Lepidoptera in Central and Northern Europe can now be largely identified to the species level [12].

According to our study, however, the situation in southern Europe is very different. Although 485 of 614 BINs could be assigned to a Linnaean name, about one-fifth could not, as 128 BINs were not assignable to species level. Unidentified species are present in practically all families, but the majority of these can be found in some groups of the so-called "microlepidoptera" (Figures 1 and S1, Table S1). Increased fractions are found, e.g., in the Autostichidae, Gelechiidae and Coleophoridae, and are on the one hand caused by the lack of reference sequences, and on the other hand probably due to additional cryptic diversity. However, the intraspecific variation of many taxa seems to be incompletely covered due

to under-represented geographical coverage of samples in BOLD. Gaytán et al. [20] found in a small group of *Quercus* herbivore moths that even a single sequence from one of the Mediterranean Peninsulas (Iberia and Italy) increased genetic divergence. Therefore, to verify that a new BIN represents an undescribed species, additional, independent evidence, e.g., from the nuclear genome or morphology, is needed.

Although many previous studies have focused on intra- and interspecific variability of DNA barcodes, we are not aware of previous works discussing intra- and inter-BIN variability. This is likely largely due to the fact that presently BOLD does not have this functionality. We observed intra-BIN variability to be on average 1.18%, while that between the closest BINs to be 4.16% on average, i.e., the inter-BIN divergences being on average about 3.5 times as large as intra-BIN divergences (see Table S1). Interestingly, within two BINs, variability exceeded 5%, which we find remarkable given that the BIN algorithm has an initial proxy for species at 2.2% [17]. This highlights that BINs are not merely defined by a fixed cut-off value but are heavily affected by the shapes of distributions of genetic variability within and between the clusters. For the same reason, with distributions being particularly narrow and sampling intensity being high, a gap of just over 1% is sufficient to assign clusters into two different BINs in our data. Compared to the previous studies on intra- and interspecific variability in lepidopteran barcodes (e.g., [12]), the BINs appear to be characterized by somewhat lower distances to nearest neighbors. The reasons for this should be studied in detail but it is likely affected by undetected cryptic diversity and other operational factors when basing the comparison on Linnean names.

Unfortunately, barcoding failed for about 10% of the material. Particularly samples of Micropterigidae did not work well with standard protocols. Based on our previous experience, DNA barcoding of species of this family generally shows a low rate of success, probably due to partial primer incompatibility. The same may be true for some other groups and species, but to our knowledge this has not been systematically studied in Lepidoptera. Furthermore, a considerable number of specimens from taxa with small body size, such as Nepticulidae and Elachistidae, likely failed because of low sample quality or quantity. Barcoding in general is unproblematic in these families, suggesting that the primer issues are not common. In our experience, the recovered 90% success rate is normal with pinned fresh samples of Lepidoptera.

The unexpectedly high number of hitherto unpublished species for the fauna of Greece largely reflects the lack of publicly available faunistic data rather than an absence of samples in private and institutional collections. We are personally aware that several of the new national records published here have been observed in the country but remained unreported. Furthermore, there is always a risk that already published faunistic data have remained unnoticed. Taxa not yet identified to species level very likely include additional records of faunistic or taxonomic interest.

The high rate (c. 20% of the inventory) of undetermined and possibly undescribed species in our sample is a significant limiting factor for faunistic-ecological and nature conservation-oriented studies. These gaps are particularly significant for possible future monitoring programs, especially for automated methods such as the use of malaise traps.

It is therefore strongly recommended to supplement the already advanced European barcode library of Lepidoptera specifically with Mediterranean taxa that have not yet been sequenced. A (near) complete DNA barcode reference library for this area would probably support taxonomic research and accelerate descriptions of new species, hence bringing us one step closer to the full bio-literacy of European lepidopteran diversity.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/d14020118/s1, Figure S1: Neighbor joining tree based on COI of specimens used in this study; Table S1: Specimens used in this study and inter- and intra-BIN distances from BOLD.

Author Contributions: Conceptualization, P.H.; Data curation, P.H.; Formal analysis, P.H. and M.M.; Methodology, P.H. and M.M.; Writing—original draft, P.H. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All 946 COI sequences are available in DS-LEPGREEC on BOLD (https://dx.doi.org/10.5883/DS-LEPGREEC) at https://www.boldsystems.org/ (accessed on 28 December 2021).

Acknowledgments: P.H. is particularly grateful to Maria Kamilari for tremendous help and support with applications. Furthermore, the BOLD staff under the leadership of Paul D.N. Hebert are thanked for their continuous support. We specifically thank Dina Soliman and Brianne St. Jacques for help with conducting calculations of BIN divergences. The Greece Ministry of Environment is acknowledged for issuing necessary permits to P.H. Robert (Bob) Heckford kindly helped with linguistic improvement of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Roslin, T.; Somervuo, P.; Pentinsaari, M.; Hebert, P.D.N.; Agda, J.; Ahlroth, P.; Anttonen, P.; Aspi, J.; Blagoev, G.; Blanco, S.; et al. A molecular-based identification resource for the arthropods of Finland. *Mol. Ecol. Res.* **2021**, *22*, 803–822. [CrossRef] [PubMed]
2. Hausmann, A.; Haszprunar, G.; Segerer, A.H.; Speidel, W.; Behounek, G.; Hebert, P.D.N. Now DNA barcoded: The butterflies and larger moths of Germany (Lepidoptera: Rhopalocera, Macroheterocera). *Spixiana* **2011**, *34*, 47–58.
3. Huemer, P.; Mutanen, M.; Sefc, K.M.; Hebert, P.D.N. Testing DNA Barcode Performance in 1000 Species of European Lepidoptera: Large Geographic Distances Have Small Genetic Impacts. *PLoS ONE* **2014**, *9*, e115774. [CrossRef] [PubMed]
4. Huemer, P.; Wiesmair, B. DNA-Barcoding der Tagfalter (Lepidoptera, Papilionoidea) Österreichs—unbekannte genetische Vielfalt im Zentrum Europas. *Wiss. Jahrb. Tirol. Landesmuseen* **2017**, *10*, 8–33.
5. Huemer, P.; Wieser, C.; Stark, W.; Hebert, P.D.N.; Wiesmair, B. DNA barcode library of megadiverse Austrian Noctuoidea (Lepidoptera)—a nearly perfect match of Linnean taxonomy. *Biodivers. Data J.* **2019**, *7*, e37734. [CrossRef] [PubMed]
6. Huemer, P.; Hebert, P.D.N. DNA Barcode Bibliothek der Schmetterlinge Südtirols und Tirols (Italien, Österreich)—Impetus für integrative Artdifferenzierung im 21. Jahrhundert. *Gredleriana* **2016**, *16*, 141–164.
7. Huemer, P.; Karsholt, O. Commented checklist of European Gelechiidae (Lepidoptera). *ZooKeys* **2020**, *921*, 65–140. [CrossRef] [PubMed]
8. Huemer, P.; Karsholt, O.; Aarvik, L.; Berggren, K.; Bidzilya, O.; Junnilainen, J.; Landry, J.-F.; Mutanen, M.; Nupponen, K.; Segerer, A.; et al. DNA barcode library for European Gelechiidae (Lepidoptera) suggests greatly underestimated species diversity. *ZooKeys* **2020**, *921*, 141–157. [CrossRef] [PubMed]
9. Lopez-Vaamonde, C.; Kirichenko, N.; Cama, A.; Doorenweerd, C.; Godfray, H.C.J.; Guiguet, A.; Gomboc, S.; Huemer, P.; Landry, J.-F.; Laštůvka, A.; et al. Evaluating DNA Barcoding for Species Identification and Discovery in European Gracillariid Moths. *Front. Ecol. Evol.* **2021**, *9*, 626752. [CrossRef]
10. Hausmann, A.; Godfray, H.C.J.; Huemer, P.; Mutanen, M.; Rougerie, R.; van Nieukerken, E.J.; Ratnasingham, S.; Hebert, P.D.N. Genetic Patterns in European Geometrid Moths Revealed by the Barcode Index Number (BIN) System. *PLoS ONE* **2013**, *8*, e84518. [CrossRef]
11. Dinca, V.; Dapporto, L.; Somervuo, P.; Vodă, R.; Cuvelier, S.; Gascoigne-Pees, M.; Huemer, P.; Mutanen, M.; Hebert, P.D.N.; Vila, R. High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. *Comm. Biol.* **2021**, *4*, 1–11. [CrossRef] [PubMed]
12. Mutanen, M.; Kivelä, S.M.; Vos, R.A.; Doorenweerd, C.; Ratnasingham, S.; Hausmann, A.; Huemer, P.; Dincă, V.; van Nieukerken, E.J.; Lopez-Vaamonde, C.; et al. Species-Level Para- and Polyphyly in DNA Barcode Gene Trees: Strong Operational Bias in European Lepidoptera. *Syst. Biol.* **2016**, *65*, 1024–1040. [CrossRef] [PubMed]
13. Ratnasingham, S.; Hebert, P.D.N. BOLD: The Barcode of Life Data System. *Mol. Ecol. Notes* **2007**, *7*, 355–364. [CrossRef] [PubMed]
14. Lepiforum: Website zur Bestimmung von Schmetterlingen (Lepidoptera) und Ihren Präimaginalstadien. Available online: http://www.lepiforum.eu/bh/downloads/2021-iii-20/Lepiforums-Europaliste_Schmetterlinge_Version_8.3_Stand_2021_03_20 (accessed on 28 December 2021).

15. De Waard, J.R.; Ivanova, N.V.; Hajibabaei, M.; Hebert, P.D.N. Assembling DNA barcodes: Analytical protocols. In *Methods in Molecular Biology: Environmental Genomics*; Martin, C.C., Ed.; Humana Press Inc.: Totowa, NJ, USA, 2008; pp. 275–293.
16. Ortiz, A.S.; Rubio, R.M.; Guerrero, J.J.; Garre, M.J.; Serrano, J.; Hebert, P.D.N.; Hausmann, A. Close congruence between Barcode Index Numbers (Bins) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. *Biodiv. Data J.* **2017**, *5*, e19840. [CrossRef] [PubMed]
17. Ratnasingham, S.; Hebert, P.D.N. A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. *PLoS ONE* **2013**, *8*, e66213. [CrossRef] [PubMed]
18. Gozmány, L. *The Lepidoptera of Greece and Cyprus. Fauna Graecia IX*; Hellenic Zoological Society: Athens, Greece, 2012; Volume 1.
19. Karsholt, O.; van Nieukerken, E.J. Lepidoptera, Moths. Fauna Europaea 2013, Version 2021.12. Available online: https://fauna-eu.org (accessed on 1 December 2021).
20. Gaytán, Á.; Bergsten, J.; Canelo, T.; Pérez-Izquierdo, C.; Santoro, M.; Bonal, R. DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots. *Ecol. Evol.* **2020**, *10*, 10754–10772. [CrossRef] [PubMed]