Biodiversity Data Journal 12: e117169
doi: 10.3897/BDJ.12.e117169

Data Paper

# The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese moths

Sónia Ferreira[‡,§,|], Martin F. V. Corley[¶], João Nunes[#], Jorge Rosete[¤], Sasha Vasconcelos[«,»], Vanessa A. Mata[‡,§], Joana Veríssimo[‡,§,^], Teresa L Silva[‡,§], Pedro Sousa[‡,˅], Rui Andrade[¦], José Manuel Grosso-Silva[ˀ], Catarina J. Pinho[‡,§,^], Cátia Chaves[‡,§], Filipa MS Martins[‡,§], Joana Pinto[‡,§], Pamela Puppo[˂], Antonio Muñoz-Mérida[‡,§], John Archer[‡,§], Joana Pauperio[¢,ℓ], Pedro Beja[‡,§,«]

‡ CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairao, Universidade do Porto, 4485-661 Vairao, Vila do Conde, Portugal
§ BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairao, 4485-661 Vairao, Vila do Conde, Portugal
| EBM, Estação Biológica de Mértola, Praça Luís de Camões, Mértola, Portugal
¶ Pucketty Farm Cottage, Faringdon, Oxfordshire SN7 8JP, U.K, Oxfordshire, United Kingdom
# Rua Eduardo Joaquim Reis Figueira, 1104 RC AR, 4440-647, Valongo, Portugal
¤ Urbanização Lourisol, Rua Manuel Cerqueira Nobrega, Lote 16, 2. frente, P-3105-165 Louriçal, Pombal, Portugal
« CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal
» Department of Ecology, Swedish University of Agricultural Sciences, PO Box 7044, 750 07, Uppsala, Sweden
^ Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4169-007, Porto, Portugal
˅ International Union for Conservation of Nature (IUCN), Species Survival Commission (SSC), Spider and Scorpion Specialist Group, Gland, Switzerland
¦ Rua Calouste Gulbenkian 237 4H3, 4050-145, Porto, Portugal
ˀ Museu de História Natural e da Ciência da Universidade do Porto, Porto, Portugal
˂ Marshall University, Department of Biological Sciences, Huntington, West Virginia, United States of America
¢ CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto,, 4485–661 Vairão, Vila do Conde, Portugal
ℓ European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom

Corresponding author: Sónia Ferreira (hiporame@gmail.com)

Academic editor: Richard Mally

Received: 11 Dec 2023 | Accepted: 19 Feb 2024 | Published: 16 May 2024

Citation: Ferreira S, Corley MFV, Nunes J, Rosete J, Vasconcelos S, Mata VA, Veríssimo J, Silva T, Sousa P, Andrade R, Grosso-Silva JM, Pinho CJ, Chaves C, Martins FM, Pinto J, Puppo P, Muñoz-Mérida A, Archer J, Pauperio J, Beja P (2024) The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese moths. Biodiversity Data Journal 12: e117169. https://doi.org/10.3897/BDJ.12.e117169

## Abstract

### Background

The InBIO Barcoding Initiative (IBI) Dataset - DS-IBILP08 contains records of 2350 specimens of moths (Lepidoptera species that do not belong to the superfamily Papilionoidea). All specimens have been morphologically identified to species or subspecies level and represent 1158 species in total. The species of this dataset correspond to about 42% of mainland Portuguese Lepidoptera species. All specimens were collected in mainland Portugal between 2001 and 2022. All DNA extracts and over 96% of the specimens are deposited in the IBI collection at CIBIO, Research Center in Biodiversity and Genetic Resources.

### New information

The authors enabled "The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese moths" in order to release the majority of data of DNA barcodes of Portuguese moths within the InBIO Barcoding Initiative. This dataset increases the knowledge on the DNA barcodes of 1158 species from Portugal belonging to 51 families. There is an increase in DNA barcodes of 205% in Portuguese specimens publicly available. The dataset includes 61 new Barcode Index Numbers. All specimens have their DNA barcodes publicly accessible through BOLD online database and the distribution data can be accessed through the Global Biodiversity Information Facility (GBIF).



## Keywords

Lepidoptera, occurrence records, species distributions, continental Portugal, DNA barcode, cytochrome c oxidase subunit I (COI)


## Introduction

The Portuguese fauna of Lepidoptera is relatively rich with 2775 species recorded so far (Corley et al. 2023). Of these, over 85 species have been described only in this century, with an increasing number of researchers contributing to the knowledge of the country's Lepidopteran fauna. During this period, 17 newly-endemic species have been described (*Phalacropterix fritschi* Hättenschwiler, 2003; *Coleophora lusitanica* Baldizzone & Corley, 2004; *Scrobipalpa corleyi* Huemer & Karsholt, 2010; *Infurcitinea corleyi* Gaedike, 2011; *Dahlica estrela* (Arnscheid, 2012); *Lobesia arzilae* Trematerra, 2014; *Willibaldiana culatrae* Trematerra, 2014; *Denisia piresi* Corley, 2014; *Filatima algarbiella* Corley, 2014; *Aroga eatoni* Corley & Goodey, 2014; *Ekboarmia miniaria* Skou, Stüning & Sihvonen, 2016; *Chrysoclista soniae* Corley, 2017; *Megacraspedus occidentellus* Huemer & Karsholt, 2018; *Afriberina salemae* Skou & Sihvonen, 2019; *Mondeguina atlanticella* Corley & Rosete,

2020; *Heterogynis cynetis* de Freina, Monasterio, Escobés, Hinojosa & Vila, 2020; *Epermenia lusitanica* Gaedike, 2022) (Corley 2015, Skou et al. 2017, Corley 2017, Huemer and Karsholt 2018, Müller et al. 2019, Corley et al. 2020, De Freina et al. 2021, Gaedike 2022) and others are known to exist (Corley, unpublished data). Since 2006, additions have been published on a yearly basis in the series "New and interesting Portuguese Lepidoptera records" (e.g. Corley et al. (2018), Corley et al. (2019b)). In the last two decades, over 500 species have been added and an annotated checklist was published in 2015 (Corley 2015). In 2021, a distribution dataset with over 50,000 records referring to specimens of 68 families, 925 genera and 2311 species/subspecies has been released in GBIF (Corley and Afonso 2021). Additionally, in 2021, a new citizen-science project created the Portuguese Moth Recording Scheme called "Rede de Estações de Borboletas Nocturnas" (REBN). This project promotes public participation in the production of faunistic data. It released a first dataset that consists of a collection of personal observations by the participants (Nunes et al. 2023) and the records produced will continue to be released at regular intervals.

In parallel with the efforts to produce and make available reliable data on the taxonomy and chorology of Portuguese moths, the IBI was created - the InBIO Barcoding Initiative, a DNA barcoding initiative by the Research Network in Biodiversity and Evolutionary Biology - InBIO as a result of the paucity of genetic data on Portuguese biodiversity. The InBIO Barcoding Initiative (IBI) makes use of High-Throughput Sequencing technologies to construct a reference collection of morphologically identified Portuguese specimens and respective DNA barcodes (e.g. Rebelo et al. 2020, Ferreira et al. 2021, Oliveira et al. 2021, Pauperio et al. 2023, Rosa et al. 2023). DNA barcoding offers a rapid, cost-effective alternative tool for both the identification of described species and the discovery of new ones (Hebert et al. 2003, Savolainen et al. 2005, Mitchell 2008, our papers). In fact, a few new species have seen their distinctness confirmed by DNA barcodes leading to their description including *Ypsolopha rhinolophi* Corley, 2019 that was found to be part of the diet of horseshoe bats in a study using DNA metabarcoding (e.g. Martin and Ferreira (2019), Corley et al. (2019a), Corley et al. (2020)). DNA barcoding provides a powerful tool by using a short fragment of DNA to assign any organism to a species in a rapid and automated way (Hebert, Cywinska, Ball and DeWaard 2003). The number of DNA barcode reference sequences available in public databases has increased remarkably and DNA barcoding has been broadly adopted (e.g. Hebert et al. 2003,Mortágua et al. 2019, Azevedo et al. 2020, da Silva et al. 2020, Fais et al. 2020, Mata et al. 2020, Velarde-Garcéz et al. 2023). However, for many groups and geographical regions, these databases are still very incomplete, which limits the general application of DNA barcoding in biodiversity research (Elbrecht et al. 2017). DNA barcodes of Lepidoptera of Iberian Peninsula are being progressivelly documented in recent years (e.g. Ortiz et al. 2017, Ortiz et al. 2023) and the present work represents the first project to generate DNA barcodes for Portuguese moths at a faunal level and, thus, represents a major step in documenting the genetic diversity of the Portuguese moth fauna.

## General description

**Purpose:** This dataset aims to provide a contribution to the knowledge on DNA barcodes of Portuguese moths. Such a library should facilitate DNA-based identification of species for both traditional molecular studies and DNA metabarcoding studies and constitute a valuable resource for taxonomic and ecological research on Lepidoptera, with the focus on the Portuguese fauna.

**Additional information:** Figs 1, 2 illustrate examples of the diversity of species that are part of the dataset of distribution data and DNA barcodes of Portuguese moths. All sequences of cytochrome c oxidase I (COI) DNA barcodes are 658 base pairs (bp) long, except for one with 422 bp (Fig. 3, Suppl. material 1). Sequences are distributed in 1199 Barcode Index Numbers (BINs), 63 being unique to this dataset. Average nucleotide composition of the sequences is 39.5% thymine (T), 15.5% cytosine (C), 30.5% adenine (A) and 14.6% guanine (G), for a total GC content of 30.1% for the COI barcode fragment analysed. Genetic p-distances ranged from 0.00% between the pair *Pleurota honorella* and *Pleurota planella* as previously found with other specimens in BIN BOLD:AEC9855; and 18.5% between *Tinea trinotella* and *Tinea murariella*. Intraspecific genetic p distances ranged from 0.00% to 11.1% in *Pleurota bicostella* group. BOLD Systems retrieved BINs to all specimens in the dataset (Ratnasingham and Hebert 2013). Only 4% of the species with more than one specimen (n = 670) had sequences assigned to two BINs (n = 31). Two species in our dataset present up to three BINs, namely *Ephestia welseriella* (BOLD:ABW1548, BOLD:AEC9178 and BOLD:AEC9179) and *Cydia fagiglandana* (BOLD:AAC5023, BOLD:ACS3074 and BOLD:AEB8145). Multiple BINs are relatively frequent in Lepidoptera species with higher percentages found in studies with higher number of specimens per species and with larger geographic scope. For example, Huemer et al. (2020) and Dincă et al. (2021) found that 20% of species of European Gelechiidae and 12.2% species of European butterflies exhibit deep splits, while studies of more limited geographical scope found 6% of species in Bavarian Geometrids (Hausmann et al. 2011) or even no deep splits in the case of Maltese islands Vella et al. (2022). The dataset includes 10 shared BINs between taxa possibly in need of taxonomical revision:

BOLD:AAA7740 (*Yponomeuta cagnagella* and *Yponomeuta evonymella*); BOLD:AAA9515 (*Chloroclysta miata* and *Chloroclysta siterata*); BOLD:AAD0839 (*Macrothylacia digramma* and *Macrothylacia rubi*); BOLD:AAB4833 (*Oligia strigilis* and *Oligia versicolor*); BOLD:AAD6780 (*Cryphia algae* and *Cryphia pallida*); BOLD:AAF0005 (*Lozotaeniodes cupressana* and *Lozotaeniodes formosana*); BOLD:ABV4113 (*Cleonymia diffluens* and *Cleonymia yvanii*); BOLD:ACE8354 (*Euxoa oranaria* and *Euxoa tritici*); BOLD:ACY5987 (*Pleurota andalusica* and *Pleurota ericella*); BOLD:AEC9855 (*Pleurota honorella* and *Pleurota planella*).

From the 2775 species belonging to 77 families recorded from continental Portugal (Corley et al. 2023), 235 species belonging to 40 families remain without DNA barcodes and 26 families have not a single DNA barcoded species (Suppl. material 6). Of these, 103

species have already specimens registered in BOLD SYSTEMS lacking DNA barcodes, while specimens of the remaining 132 species are still needed.

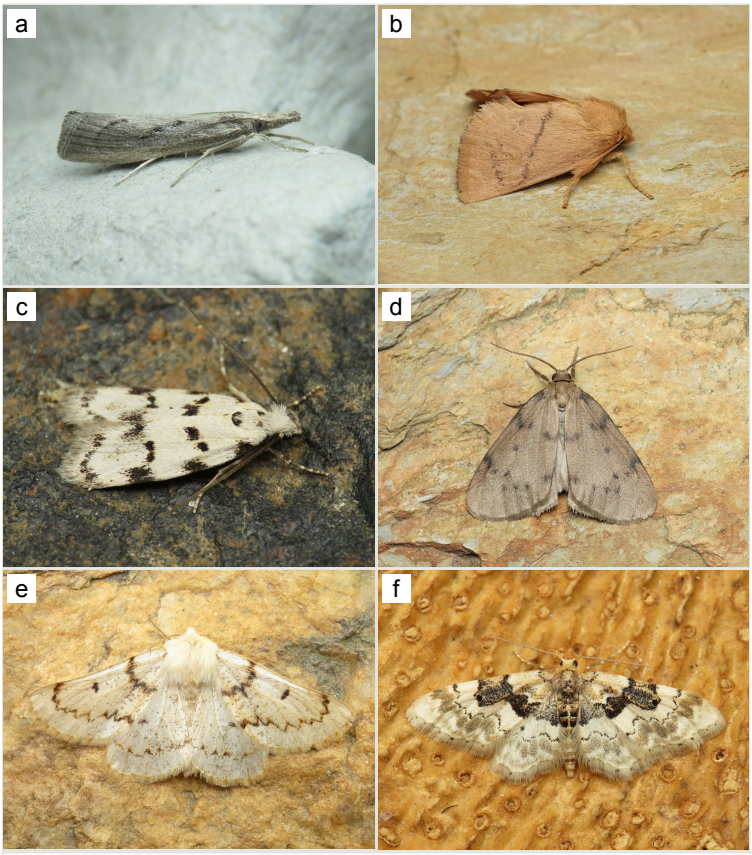


Figure 1.

Examples of the diversity of species that are part of the dataset of distribution data and DNA barcodes of Portuguese moths. All photos by João Nunes.

**a**: *Agriphila tersellus* Lederer, 1855, Crambidae - BIN URI BOLD:AEC8773; doi
**b**: *Hoyosia codeti* (Oberthür, 1883), Limacodidae - BIN URI BOLD:ADV3396; doi
**c**: *Symmoca serrata* Gozmany, 1985, Autostichidae - BIN URI BOLD:AEO2252; doi
**d**: *Paidia rica* Freyer, 1855, Erebidae - BIN URI BOLD:AEC7499; doi
**e**: *Dyscia distinctaria* (Bang-Haas, 1910), Geometridae - BIN URI BOLD:ACF1137; doi
**f**: *Idaea figuraria* (Bang-Haas, 1907), Geometridae - BIN URI BOLD:AEC6951. doi

## Project description

**Title:** The name "The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese moths" refers to the first data release of DNA barcodes and distribution data of Portuguese moths within the InBIO Barcoding Initiative.

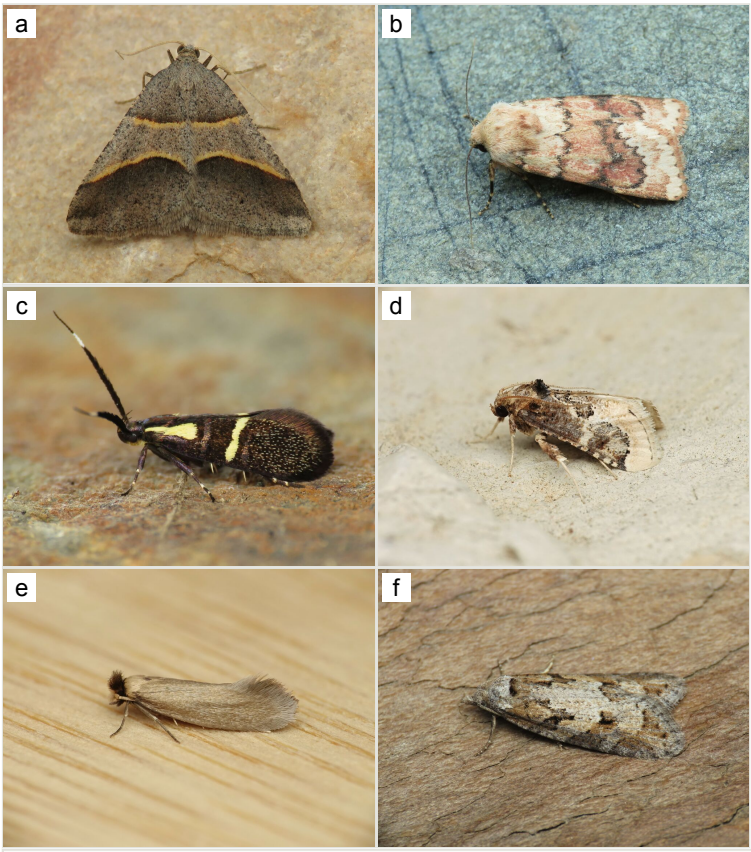


Figure 2.

Examples of the diversity of species that are part of the dataset of distribution data and DNA barcodes of Portuguese moths (cont.). All photos by João Nunes.

**a**: *Perigune convergata* (Villers, 1789), Geometridae - BIN URI BOLD:AEH2380; doi
**b**: *Dichagyris constanti* Millière, 1860, Noctuidae - BIN URI BOLD:AED6734; doi
**c**: *Dasycera oliviella* (Fabricius, 1794), Oecophoridae - BIN URI BOLD:ADT8549; doi
**d**: *Loryma egregialis* (Herrich-Schäffer, 1838), Pyralidae - BIN URI BOLD:AEC8952; doi
**e**: *Crassicornella agenjoi* (Petersen, 1957), Tineidae - BIN URI BOLD:ADT5404; doi
**f**: *Eana nervana* (de Joannis, 1908), Tortricidae - BIN URI BOLD:ADS8408. doi

**Personnel:** Pedro Beja (project coordinator), Sónia Ferreira (taxonomist and IBI manager), Martin F.V. Corley (taxonomist, lepidopterist), Cátia Chaves (project technician), Filipa M.S. Martins (molecular biologist), Vanessa A. Mata (molecular biologist), Antonio Muñoz-Mérida (bioinformatician), John Archer (bioinformatician), Joana Paupério (molecular biologist), Catarina J. Pinho (project technician), Joana C. Pinto (project technician), Pamela Puppo (molecular biologist), Teresa L Silva (molecular biologist), Pedro Sousa (project technician), Sasha Vasconcelos (contributor), Joana Veríssimo (molecular biologist), all affiliated to CIBIO-InBIO, University of Porto, José Manuel Grosso-Silva (entomologist), affiliated to the MHNC-UP, University of Porto and Rui Andrade

(entomologist), João Nunes (lepidopterologist), Jorge Rosete (lepidopterist), independent researchers.

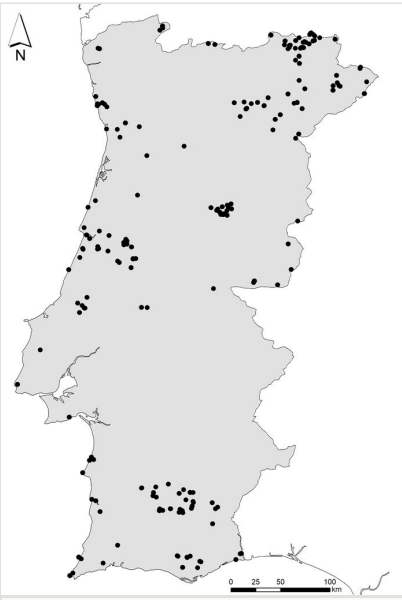


Figure 3. doi

Map of the localities where DNA barcoded Lepidoptera samples were collected in continental Portugal.

**Study area description:** Continental Portugal (Fig. 3).

**Design description:** Lepidoptera specimens were collected in the field, morphologically identified and DNA barcoded.

**Funding:** The present work was funded by National Funds through FCT-Fundação para a Ciência e a Tecnologia in the scope of the project LA/P/0048/2020. InBIO Barcoding Initiative was funded by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 668981 and by the project PORBIOTA – Portuguese E-Infrastructure for Information and Research on Biodiversity (POCI-01-0145-FEDER-022127), supported by Operational Thematic Program for Competitiveness and Internationalization (POCI), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). The work was partially Funded by Horizon Europe under the Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00173; and by the UK Research and Innovation under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. The fieldwork benefitted from EDP Biodiversity Chair, the project "Promoção dos serviços deecossistemas no Parque Natural Regional do Vale do Tua: Controlo de Pragas Agrícolas eFlorestais por Morcegos" funded by the Agência de Desenvolvimento Regional

do Vale doTua and includes research conducted at the Long Term Research Site of Baixo Sabor (LTER_EU_PT_002). SF and VM were funded by the FCT through the programme 'Stimulus of Scientific Employment, Individual Support—3rd Edition' (https://doi.org/10.54499/2020.03526.CEECIND/CP1601/CT0010; https://doi.org/10.54499/2020.02547.CEECIND/CP1601/CT0006). CJP, JV and FMSM by PhD grants (SFRH/BD/145851/2019; SFRH/BD/133159/2017; SFRH/BD/104703/2014) funded by FCT.

## Sampling methods

**Description:** Continental Portugal (Fig. 3).

**Sampling description:** Specimens were collected during field expeditions throughout continental Portugal, from 2001 to 2020. They were captured at night using light traps, the latter with UV LEDs, mixed light or mercury vapour lamps or during the day by direct search. All specimens were observed in the field and, in most cases, they could be readily identified to species level by an experienced taxonomist (Martin Corley). Such specimens were preserved in 96% ethanol and stored at the InBIO Barcoding Initiative reference collection (Vairão, Portugal), where they can be re-examined and genitalia dissected, if needed. Specimens that could not be identified in the field (n = 84) were pinned and dried for subsequent examination in the laboratory. They were then stored at the Research Collection of Martin Corley or the Private Collections of Jorge Rosete or J.M. Grosso-Silva.

DNA extraction was performed using either the 96-Well Plate Animal Genomic DNA Mini-Preps Kit (Bio Basic, Ontario, Canada) or the QIAamp DNA Micro Kit (Qiagen, Germany) which is designed to extract higher concentrations of genetic material from samples with small amounts of DNA. Amplification was performed using two different primer pairs that amplify partially overlapping fragments (LC + BH) of the 658 bp barcoding region of the COI mitochondrial gene. We used the primers FwhF1 (Vamos et al. 2017) + C_R ( Shokralla et al. 2015) for LC and BF3 (Elbrecht et al. 2019) + BR2 (Elbrecht and Leese 2017) for BH amplification, all modified with 5' adaptors sequences to be compatible with a two-step protocol. PCRs were performed in 10 µl reactions, containing 5 µl of Multiplex PCR Master Mix (Qiagen, Germany), 0.3 µl of each 10 mM primer and 1-2 µl of DNA, with the remaining volume in water. PCR cycling conditions consisted of an initial denaturation at 95ºC for 15 min, followed by 45 cycles of denaturation at 95ºC for 30 sec, annealing at 45ºC and 50ºC, respectively, for 45 sec and extension at 72ºC for 45 sec and a final elongation step at 60ºC for 10 min.

Successful amplification was validated through 2% agarose gel electrophoresis stained with GelRed (Biotium, USA) and samples selected for sequencing proceeded for a second-round PCR where Illumina P5 and P7 adapters with custom 7 bp long barcodes were attached to each first PCR product. The second PCR was performed in a volume of 10 µl, including 5 µl of KAPA HiFi PCR Kit (KAPA Biosystems, Cape Town, South Africa), 0.5 µl of each 10 mM indexing primer and 2 µl of diluted first PCR product (usually 1:4). PCR cycling conditions were as follows: initial denaturation at 95ºC for 3 min, with 8-10 cycles (adjusted to sample quality) of denaturation at 95ºC for 30 sec, annealing at 50ºC for 60

sec and extension at 72ºC for 45 sec and a final elongation step at 60ºC for 10 min. The amplicons were purified using AMPure XP beads (Beckman Coulter, U.S.A.) and quantified using NanoDrop 1000 (Thermo Scientific, U.S.A.). Clean PCR products were then pooled equimolarly per fragment. Each pool was quantified with KAPA Library Quantification Kit Illumina Platforms (KAPA Biosystems, Cape Town, South Africa) and the 2200 Tapestation System (Agilent Technologies, California, USA) was used for fragment length analysis prior to sequencing (Paupério et al. 2018). DNA sequencing was done at CIBIO facilities on an Illumina MiSeq benchtop system, using V2 MiSeq sequencing kits (2 x 250 bp) (Illumina, California, U.S.A.).

Illumina sequencing reads were processed using OBITools (Boyer et al. 2015) and VSEARCH (Rognes et al. 2016). Briefly, paired-end reads were aligned, collapsed into exact sequence variants, filtered by length, denoised and checked for chimeras. The resulting sequences from both LC and BH fragments of each sample were further assembled using CAP3 (Huang and Madan 1999) to produce a single 658 bp contig per sample.

**Quality control:** All DNA barcodes sequences were compared against the BOLD database and the 99 top hits were inspected in order to detect possible issues due to contaminations or misidentifications.

**Step description:** 1. Specimens were collected in 234 different localities. Fieldwork was carried out between 2001 and 2022.

2. Selected specimens were pinned and dried and are preserved in three private collections. Otherwise, specimens collected as tissue samples were stored in 96% ethanol in the IBI collection at CIBIO, Research Center in Biodiversity and Genetic Resources (Vairão, Portugal).

3. All specimens were morphologically identified and DNA barcoded. To sequence the 658 bp COI DNA barcode fragment, one leg was removed from each individual, DNA was extracted and then amplified. All DNA extracts were deposited in the IBI collection.

4. All sequences in the dataset were submitted to BOLD and GenBank databases and, to each sequenced specimen, the morphological identification was contrasted with the results of the BLAST of the newly-generated DNA barcodes in the BOLD Identification Engine.

## Geographic coverage

**Description:** Continental Portugal

**Coordinates:** 36.960 and 42.124 Latitude; -9.467 and -6.229 Longitude.

## Taxonomic coverage

**Description:** This dataset is composed of data relating to 2364 Lepidoptera specimens. All specimens were determined to species level. Overall, 1170 species are represented in the dataset. These species belong to 51 families and 598 genera (Suppl. material 1). The dataset is represented mostly by seven families that include more than 75% of the specimens, of which Noctuidae, Geometridae and Tortricidae account for more than 20% of each of them (Fig. 4). Twenty-six families known to be present in Portugal are not represented in the dataset. The *Idaea* genus accounts for 4% of the total collected specimens and five other genera accounts for 1 - 3% (Fig. 5).

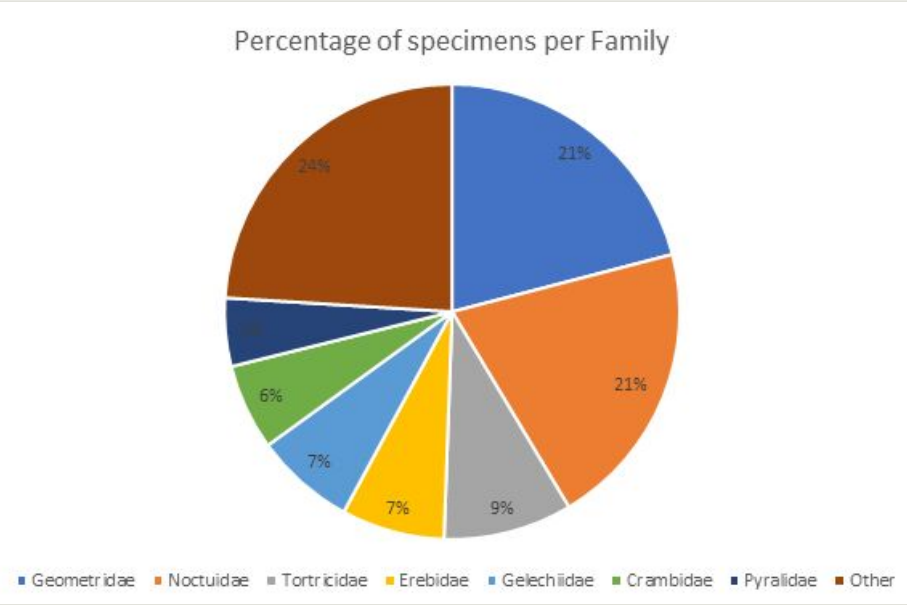


Figure 4. doi

Distribution of specimens, in percentage, per moth family present in the dataset. Families representing less than 3% of the total specimens are represented together in the respective graph.

## Temporal coverage

**Notes:** The sampled material was collected in the period from 2001 to 2020.

## Collection data

**Collection name:** InBIO Barcoding Initiative

**Collection identifier:** 4ec2b246-f5fa-4b90-9a8d-ddafc2a3f970

**Specimen preservation method:** "Dry", "Alcohol"

**Curatorial unit:** Dry voucher - 1- 84, Voucher tube - 1 to 2286, DNA extractions - 1 to 2364

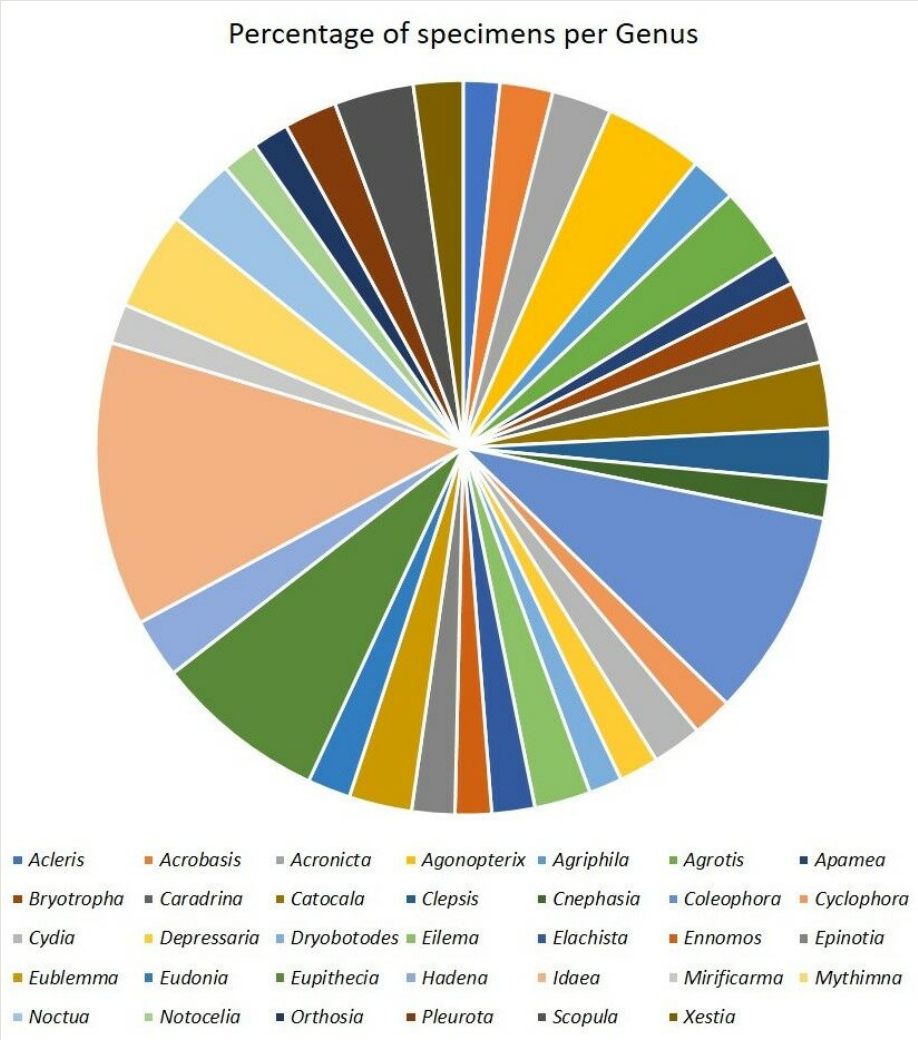


Figure 5. doi

Distribution of specimens, in percentage, per moth genus present in the dataset. Genera represented by less than 0.4% (of total specimens) are not represented in the graph and correspond to 71% of genera in the dataset.

## Usage licence

## Data resources

**Data package title:** The InBIO Barcoding Initiative Database: contribution to the knowledge on DNA barcodes of moths (Lepidoptera)

**Resource link:** http://dx.doi.org/10.5883/DS-IBILP08

**Number of data sets:** 1

**Data set name:** IBI-Lepidoptera 08

**Data format:** dwc, xml, tsv, fasta

**Description:** The InBIO Barcoding Initiative Database: contribution to the knowledge on DNA barcodes of Portuguese moths Lepidoptera dataset can be downloaded from the Public Data Portal of BOLD (http://www.boldsystems.org/index.php/Public_SearchTerms?query=DS-IBILP08) in different formats (data as dwc, xml or tsv and sequences as fasta files). Alternatively, BOLD users can log-in and access the dataset via the Workbench platform of BOLD. All records are also searchable within BOLD, using the search function of the database.

The version of the dataset, at the time of writing the manuscript, is included as Suppl. materials 2, 3, 4, 5 in the form of two text files with specimen data information and one fasta file containing all sequences as downloaded from BOLD.

| Column label | Column description |
|---|---|
| processid | Unique identifier for the sample. |
| sampleid | Identifier for the sample being sequenced, i.e. IBI catalogue number at Cibio&#2;InBIO,Porto University. Often identical to the "Field ID" or "Museum. |
| recordID | Identifier for specimen assigned in the field. |
| catalognum | Catalogue number. |
| fieldnum | Field number. |
| institution_storing | The full name of the institution that has physical possession of the voucher specimen. |
| bin_uri | Barcode Index Number system identifier. |
| phylum_taxID | Phylum taxonomic numeric code. |
| phylum_name | Phylum name. |
| class_taxID | Class taxonomic numeric code. |
| class_name | Class name. |
| order_taxID | Order taxonomic numeric code. |
| order_name | Order name. |

| | |
|---|---|
| family_taxID | Family taxonomic numeric code. |
| family_name | Family name. |
| subfamily_taxID | Subfamily taxonomic numeric code. |
| subfamily_name | Subfamily name. |
| genus_taxID | Genus taxonomic numeric code. |
| genus_name | Genus name. |
| species_taxID | Species taxonomic numeric code. |
| species_name | Species name. |
| identification_provided_by | Full name of primary individual who assigned the specimen to a taxonomic group. |
| identification_method | The method used to identify the specimen. |
| voucher_status | Status of the specimen in an accessioning process (BOLD controlled vocabulary). |
| tissue_type | A brief description of the type of tissue or material analysed. |
| collectors | The full or abbreviated names of the individuals or team responsible for collecting the sample in the field. |
| lifestage | The age class or life stage of the specimen at the time of sampling. |
| sex | The sex of the specimen. |
| lat | The geographical latitude (in decimal degrees) of the geographic centre of a location. |
| lon | The geographical longitude (in decimal degrees) of the geographic centre of a location. |
| elev | Elevation of sampling site (in metres above sea level). |
| country | The full, unabbreviated name of the country where the organism was collected. |
| province_state | The full, unabbreviated name of the province where the organism was collected. |
| region | The full, unabbreviated name of the municipality where the organism was collected. |
| exactsite | Additional name/text description regarding the exact location of the collection site relative to a geographic relevant landmark. |

## Acknowledgements

The present work was funded by National Funds through FCT-Fundação para a Ciência e a Tecnologia in the scope of the project LA/P/0048/2020. InBIO Barcoding Initiative was funded by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 668981 and by the project PORBIOTA – Portuguese E-Infrastructure for Information and Research on Biodiversity (POCI-01-0145- FEDER-022127), supported by Operational Thematic Program for Competitiveness and Internationalization (POCI), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). The work was partially Funded by Horizon Europe under the

Biodiversity, Circular Economy and Environment call (REA.B.3); co-funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00173; and by the UK Research and Innovation under the Department for Business, Energy and Industrial Strategy's Horizon Europe Guarantee Scheme. The fieldwork benefitted from EDP Biodiversity Chair, the project "Promoção dos serviços deecossistemas no Parque Natural Regional do Vale do Tua: Controlo de Pragas Agrícolas eFlorestais por Morcegos" funded by the Agência de Desenvolvimento Regional do Vale doTua and includes research conducted at the Long Term Research Site of Baixo Sabor (LTER_EU_PT_002). SF and VM were funded by the FCT through the programme 'Stimulus of Scientific Employment, Individual Support—3rd Edition' (https://doi.org/10.54499/2020.03526.CEECIND/CP1601/CT0010; https://doi.org/10.54499/2020.02547.CEECIND/CP1601/CT0006). CJP, JV and FMSM by PhD grants (SFRH/BD/145851/2019; SFRH/BD/133159/2017; SFRH/BD/104703/2014) funded by FCT. The authors would like to acknowledge all colleagues that have joined and helped in fruitful fieldwork throughout the years and shared their thoughts in many profitable discussions on Portuguese moths, namely Ana Rita Gonçalves, Ana Valadares, Carla Gomes, Daniel Oliveira, Luis Silva, Marisa Rodrigues, Pedro Lopes and Rebecca Mateus.

## References


- Azevedo J, Antunes J, Machado A, Vasconcelos V, Leão P, Froufe E (2020) Monitoring of biofouling communities in a Portuguese port using a combined morphological and metabarcoding approach. Scientific Reports 10 (1). https://doi.org/10.1038/s41598-020-70307-4
- Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E (2015) OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. Molecular Ecology Resources 16 (1): 176-182. https://doi.org/10.1111/1755-0998.12428
- Corley MFV (2015) Lepidoptera of Continental Portugal. A fully revised list. 1 Edition. Martin Corley, Faringdon, 288 pp. [In English]. [ISBN 9780993316005]
- Corley MFV (2017) *Chrysoclista soniae* (Lepidoptera, Elachistidae, Parametriotinae), a new species from Portugal. Entomologist's Gazette 68: 130-134. [In English].
- Corley MFV, Ferreira S, Grundy D, Nunes J, Pires P, Rosete J (2018) New and interesting Portuguese Lepidoptera records from 2017 (Insecta: Lepidoptera). SHILAP Revista de lepidopterología 46 (184): 551-576. https://doi.org/10.57065/shilap.747
- Corley MFV, Ferreira S, Mata V (2019a) *Ypsolopha rhinolophi* sp. nov. (Lepidoptera: Ypsolophidae), a new species from Portugal and France unveiled by bats. Zootaxa 4609 (3). https://doi.org/10.11646/zootaxa.4609.3.10
- Corley MFV, Nunes J, Rosete J, Ferreira S (2019b) New and interesting Portuguese Lepidoptera records from 2018 (Insecta: Lepidoptera). SHILAP Revista de lepidopterología 47 (188): 611-630. https://doi.org/10.57065/shilap.490
- Corley MFV, Rosete J, Ferreira S (2020) Mondeguina, a new genus for *Apatetris mediterranella* Nel & Varenne, 2012, with description of a new species from Portugal (Lepidoptera, Gelechiidae). Nota Lepidopterologica 43: 151-166. https://doi.org/10.3897/nl.43.50430

- Corley MFV, Afonso B (2021) Portuguese Lepidoptera records compiled by Martin Corley's database. 1.5. CIBIO (Research Center in Biodiversity and Genetic Resources. Release date: 2021-12-07. URL: https://doi.org/10.15468/ca4xt8
- Corley MFV, Nunes J, Rosete J, Teixeira J, Valadares A (2023) New and interesting Portuguese Lepidoptera records from 2022 (Insecta: Lepidoptera). SHILAP Revista de lepidopterología 51 (204): 609-627. https://doi.org/10.57065/shilap.784
- da Silva L, Mata V, Lopes P, Lopes R, Beja P (2020) High-resolution multi-marker DNA metabarcoding reveals sexual dietary differentiation in a bird with minor dimorphism. Ecology and Evolution 10 (19): 10364-10373. https://doi.org/10.1002/ece3.6687
- De Freina J, Monasterio León Y, Escobés Jiménez R, Vila R, Hinojosa JC (2021) *Heterogynis* Rambur, 1837 (Lepidoptera: Zygaenoidea, Heterogynidae) from Andalusia and the Algarve (Spain and Portugal): a taxonomic and faunistic review. Boletin de la Sociedad Entomológica Aragonesa 64: 75-85. URL: http://zoobank.org/842fab2f-fa0c-4dba-a95b-2f0a94184739
- Dincă V, Dapporto L, Somervuo P, Vodă R, Cuvelier S, Gascoigne-Pees M, Huemer P, Mutanen M, Hebert PN, Vila R (2021) High resolution DNA barcode library for European butterflies reveals continental patterns of mitochondrial genetic diversity. Communications Biology 4 (1). https://doi.org/10.1038/s42003-021-01834-7
- Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science 5 https://doi.org/10.7287/peerj.preprints.2044v5
- Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution 8 (10): 1265-1275. https://doi.org/10.1111/2041-210x.12789
- Elbrecht V, Braukmann TA, Ivanova N, Prosser SJ, Hajibabaei M, Wright M, Zakharov E, Hebert PN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ 7 https://doi.org/10.7717/peerj.7745
- Fais M, Bellisario B, Duarte S, Vieira PE, Sousa R, Canchaya C, Costa FO (2020) Meiofauna metabarcoding in Lima estuary (Portugal) suggests high taxon replacement within a background of network stability. Regional Studies in Marine Science 38 https://doi.org/10.1016/j.rsma.2020.101341
- Ferreira S, Oosterbroek P, Starý J, Sousa P, Mata V, da Silva L, Paupério J, Beja P (2021) The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese Diptera 02 - Limoniidae, Pediciidae and Tipulidae. Biodiversity Data Journal 9 https://doi.org/10.3897/bdj.9.e69841
- Gaedike R (2022) Two new species of the genus Epermenia Hübner, [1825] and some new distributional and taxonomic records (Lepidoptera: Epermeniidae). SHILAP Revista de lepidopterología 50 (200): 755-761. https://doi.org/10.57065/shilap.269
- Hausmann A, Haszprunar G, Hebert PN (2011) DNA Barcoding the Geometrid Fauna of Bavaria (Lepidoptera): Successes, Surprises, and Questions. PLoS ONE 6 (2). https://doi.org/10.1371/journal.pone.0017134
- Hebert PN, Cywinska A, Ball S, deWaard J (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences 270 (1512): 313-321. https://doi.org/10.1098/rspb.2002.2218
- Huang X, Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Research 9 (9): 868-877. https://doi.org/10.1101/gr.9.9.868

- Huemer P, Karsholt O (2018) Revision of the genus *Megacraspedus* Zeller, 1839, a challenging taxonomic tightrope of species delimitation (Lepidoptera, Gelechiidae). ZooKeys 800: 1-278. https://doi.org/10.3897/zookeys.800.26292
- Huemer P, Karsholt O, Aarvik L, Berggren K, Bidzilya O, Junnilainen J, Landry J, Mutanen M, Nupponen K, Segerer A, Šumpich J, Wieser C, Wiesmair B, Hebert PN (2020) DNA barcode library for European Gelechiidae (Lepidoptera) suggests greatly underestimated species diversity. ZooKeys 921: 141-157. https://doi.org/10.3897/zookeys.921.49199
- Martin CFV, Ferreira S (2019) A taxonomic revision of the Western Palaearctic genus *Cacochroa* Heinemann, 1870 (Lepidoptera, Depressariidae, Cryptolechiinae) with description of a new genus and a new species. Zootaxa 4683 (2). https://doi.org/10.11646/zootaxa.4683.2.2
- Mata V, Ferreira S, Campos R, da Silva L, Veríssimo J, Corley MV, Beja P (2020) Efficient assessment of nocturnal flying insect communities by combining automatic light traps and DNA metabarcoding. Environmental DNA 3 (2): 398-408. https://doi.org/10.1002/edn3.125
- Mitchell A (2008) DNA barcoding demystified. Australian Journal of Entomology 47 (3): 169-173. https://doi.org/10.1111/j.1440-6055.2008.00645.x
- Mortágua A, Vasselon V, Oliveira R, Elias C, Chardon C, Bouchez A, Rimet F, João Feio M, F.P. Almeida S (2019) Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecological Indicators 106 https://doi.org/10.1016/j.ecolind.2019.105470
- Müller B, Erlacher S, Hausmann A, Rajaei H, Sihvonen P, Skou P (2019) Ennominae II. In: Hausmann A, Sihvonen P, Rajaei H, Skou P (Eds) The geometrid moths of Europe. Vol. 6. Brill, Leiden, 906 pp.
- Nunes J, Cardoso H, Valadares A, Banza P, Tomás J, Valkenburg T (2023) Portuguese Moth Recording Scheme. 1.3. GBIF Portugal. URL: https://doi.org/10.15468/mbpxnd
- Oliveira D, Chaves C, Pinto J, Paupério J, Fonseca N, Beja P, Ferreira S (2021) DNA barcoding of Portuguese lacewings (Neuroptera) and snakeflies (Raphidioptera) (Insecta, Neuropterida). ZooKeys 1054: 67-84. https://doi.org/10.3897/zookeys.1054.64608
- Ortiz A, Rubio R, Guerrero J, Garre M, Serrano J, Hebert P, Hausmann A (2017) Close congruence between Barcode Index Numbers (bins) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. Biodiversity Data Journal 5 https://doi.org/10.3897/bdj.5.e19840
- Ortiz A, Rubio R, de Freina J, Guerrero J, Garre M, Yela JL (2023) DNA barcoding and morphology reveal European and western Asian Arctia villica (Linnaeus, 1758) as a complex of species (Lepidoptera, Erebidae, Arctiinae). ZooKeys 1159: 69-86. https://doi.org/10.3897/zookeys.1159.95225
- Pauperio J, Gonzalez LM, Martinez J, González M, Martins FM, Veríssimo J, Puppo P, Pinto J, Chaves C, Pinho C, Grosso-Silva JM, Quaglietta L, Silva TL, Sousa P, Alves P, Fonseca N, Beja P, Ferreira S (2023) The InBIO barcoding initiative database: DNA barcodes of Iberian Trichoptera, documenting biodiversity for freshwater biomonitoring in a Mediterranean hotspot. Biodiversity Data Journal 11 https://doi.org/10.3897/bdj.11.e97484

- Ratnasingham S, Hebert PN (2013) A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE 8 (7). https://doi.org/10.1371/journal.pone.0066213
- Rebelo H, Ferreira S, Amorim F, Horta P, Raposeira H, Santos H, Beja P, Mata V (2020) Hidden in our pockets: building of a DNA barcode library unveils the first record of *Myotis alcathoe* for Portugal. Biodiversity Data Journal 8 https://doi.org/10.3897/bdj.8.e54479
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4 https://doi.org/10.7717/peerj.2584
- Rosa P, Wood T, Silva TL, Veríssimo J, Mata V, Michez D, Beja P, Ferreira S (2023) The InBIO Barcoding Initiative Database: contribution to the knowledge on DNA barcodes of cuckoo wasps, with the description of new species from the Iberian Peninsula (Hymenoptera, Chrysididae). Biodiversity Data Journal 11 https://doi.org/10.3897/bdj.11.e98743
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1462): 1805-1811. https://doi.org/10.1098/rstb.2005.1730
- Shokralla S, Porter T, Gibson J, Dobosz R, Janzen D, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Scientific Reports 5 (1). https://doi.org/10.1038/srep09687
- Skou P, Stüning D, Sihvonen P (2017) Revision of the West-Mediterranean geometrid genus *Ekboarmia*, with description of a new species from Portugal (Lepidoptera, Geometridae, Ennominae). Nota Lepidopterologica 40 (1): 39-63. https://doi.org/10.3897/nl.40.10440
- Vamos E, Elbrecht V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. Metabarcoding and Metagenomics 1 https://doi.org/10.3897/mbmg.1.14625
- Velarde-Garcéz D, Mata V, Beja P, da Silva L (2023) DNA metabarcoding, diversity partitioning and null models reveal mechanisms of seasonal trophic specialization in a Mediterranean warbler. Molecular Ecology https://doi.org/10.1111/mec.17245
- Vella A, Mifsud CM, Magro D, Vella N (2022) DNA Barcoding of Lepidoptera Species from the Maltese Islands: New and Additional Records, with an Insight into Endemic Diversity. Diversity 14 (12). https://doi.org/10.3390/d14121090

# Supplementary materials

## Suppl. material 1: List of species that were collected and DNA barcoded within The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese moths

doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel Grosso Silva, Catarina J. Pinho, Cátia Chaves, Filipa MS Martins, Joana Pinto, Pamela Puppo, Antonio Muñoz-Mérida, John Archer, Joana Pauperio, Pedro Beja

**Data type:** occurrences

**Brief description:** List of species that were collected and DNA barcoded within this project including the sample code (IBI code), the Process ID (BOLDcode), the BIN URI (BOLD BIN) and GenBank acession number (GenBank). * Indicate species with new BINs.

Download file (212.71 kb)

## Suppl. material 2: IBI - Lepidoptera 08 library - Specimen details

doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel Grosso Silva, Joana Pauperio, Pedro Beja

**Data type:** Specimen data records

**Brief description:** The file includes information about all records in BOLD for the IBI - Lepidoptera 08 library. It contains collecting and identification data. The data are as downloaded from BOLD in the tsv format, without further processing.

Download file (924.30 kb)

## Suppl. material 3: IBI - Lepidoptera 08 - Specimen details - Darwin Core Standard

doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel Grosso Silva, Joana Pauperio, Pedro Beja

**Data type:** Specimen data records in the Darwin Core Standard format

**Brief description:** The file includes information about all records in BOLD for the IBI - Lepidoptera 08 library. It contains collecting and identification data. The data are as downloaded from BOLD in the Darwin Core Standard format, without further processing.

Download file (1.08 MB)

### Suppl. material 4: IBI- Lepidoptera 08 library - DNA sequences doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel GrossoSilva, Catarina J. Pinho, Cátia Chaves, Filipa MS Martins, Joana Pinto, Pamela Puppo, Antonio Muñoz-Mérida, John Archer, Joana Pauperio, Pedro Beja
**Data type:** Specimen genomic data, DNA sequences
**Brief description:** COI sequences in fasta format. Each sequence is identified by the BOLDProcessID, species name, genetic marker name and GenBank accession number, all separated by a vertical bar. The data are as downloaded from BOLD.
Download file (1.59 MB)

### Suppl. material 5: Phylogenetic tree (NJ) of all DNA barcodes used in the study generated in BOLD Systems doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel Grosso Silva, Catarina J. Pinho, Cátia Chaves, Filipa MS Martins, Joana Pinto, Pamela Puppo, Antonio Muñoz-Mérida, John Archer, Joana Pauperio, Pedro Beja
**Data type:** Phylogenetic tree
**Brief description:** Phylogenetic tree (NJ) of all the specimens DNA barcodes within DS-IBILP08: the IBI-Lepidoptera 08 dataset collected in continental Portugal, all of which have been morphologically identified to species level.
Download file (100.77 kb)

### Suppl. material 6: Specimens with occurence in continental Portugal missing DNA barcodes doi

**Authors:** Sónia Ferreira, Martin F. V. Corley, João Nunes, Jorge Rosete, Sasha Vasconcelos, Vanessa A. Mata, Teresa L Silva, Pedro Sousa, Rui Andrade, José Manuel Grosso Silva, Catarina J. Pinho, Cátia Chaves, Filipa MS Martins, Joana Pinto, Pamela Puppo, Antonio Muñoz-Mérida, John Archer, Joana Pauperio, Pedro Beja
**Data type:** Gap Checklist
**Brief description:** Specimens with occurrence in continental Portugal missing DNA barcodes. To each species is discriminated if the species has specimens in BOLD, but still has no DNA barcode or specimens are still needed.
Download file (14.92 kb)