

IberArthro: A free online database compiling taxonomic and distributional data on Ibero-Balearic arthropods

Saray Mañas-Jordá¹ , Raúl Acosta^{2,3} , Arturo Ariño⁴ , Enrique Baquero⁴ , Nuria Bonada^{2,5} , David Galicia⁴ , Enrique García-Barros⁶ , Antonio J. García-Meseguer¹ , Emilio García-Roselló⁷ , Marcos González⁸ , Jorge M. Lobo⁹ , Manuel Jesús López-Rodríguez¹⁰ , Jesús Martínez¹¹ , Andrés Millán¹ , Víctor J. Monserrat¹² , Miguel L. Munguira⁶ , Carlos E. Prieto¹³ , Helena Romo⁶ , Carlota Sánchez-Campaña^{2,14} , José M. Tierno de Figueroa¹⁵ , José Luis Yela¹⁶ , David Sánchez-Fernández^{1,*} 

- (1) Departament of Ecology and Hidrology, University of Murcia, Murcia, Spain.
- (2) FEHM-Lab (Freshwater Ecology, Hydrology and Management), Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain.
- (3) Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain.
- (4) Institute for Biodiversity and Environment BIOMA, University of Navarra, Pamplona, Spain.
- (5) Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.
- (6) Department of Biology, Autonomous University of Madrid (UAM) and Center for Biodiversity and Global Change Research (CIBC-UAM), Madrid, Spain.
- (7) Department of Computer Science, University of Vigo, Vigo, Spain.
- (8) Department of Zoology, Genetics and Physical Anthropology, University of Santiago de Compostela, Santiago de Compostela, Spain.
- (9) Department of Biogeography and Global Change, National Museum of Natural Sciences (MNCN-CSIC), Madrid, Spain.
- (10) Department of Ecology, University of Granada, Granada, Spain.
- (11) Department of Zoology, Genetics and Physical Anthropology, University of Santiago de Compostela, Santiago de Compostela, Spain.
- (12) Department of Biodiversity, Ecology and Evolution, Complutense University, Madrid, Spain.
- (13) Department of Zoology and Animal Cell Biology, University of the Basque Country, Bilbao, Spain.
- (14) Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona (UB), Barcelona, Spain.
- (15) Department of Zoology, University of Granada, Granada, Spain.
- (16) Department of Environmental Sciences, University of Castilla-La Mancha, Toledo, Spain.

* Corresponding author / Autor para correspondencia: David Sánchez-Fernández [davidsan@um.es]

> Received / Recibido: 03/02/2025 – Accepted / Aceptado: 04/04/2025

How to cite / Cómo citar: Mañas-Jordá, S., Acosta, R., Ariño, A., Baquero, E., Bartomeus, I., Bonada, N., Galicia, D., García-Barros, E., García-Meseguer, A.J., García-Roselló, E., González, M., Lobo, J.M., López-Rodríguez, M.J., Martínez, J., Millán, A., Monserrat, V.J., Munguira, M.L., Prieto, C.E., Romo, H., Sánchez-Campaña, C., Tierno de Figueroa, J.M., Yela, J.L., Sánchez Fernández, D. 2025. IberArthro: A free online database compiling taxonomic and distributional data on Ibero-Balearic arthropods. *Ecosistemas* 34(2): 2955. <https://doi.org/10.7818/ECOS.2955>

IberArthro: A free online database compiling taxonomic and distributional data on Ibero-Balearic arthropods

Abstract: Arthropods constitute the bulk of global biodiversity, being the most diverse metazoan phylum on the planet. Despite the relevance of arthropods, our knowledge about them remains limited, partly due to the scarcity of reliable faunistic data. The Iberian Peninsula is part of one of the global biodiversity hotspots, the Mediterranean Basin. However, very little is known on the ecology and spatial distribution of many Iberian arthropod species. We provide a database on the distribution of Ibero-Balearic arthropods (the IberArthro database) belonging to 10 taxonomic groups: Coleoptera (Aquatic), Coleoptera (Scarabaeoidea), Collembola, Hymenoptera (Apoidea), Lepidoptera (Noctuoidea), Lepidoptera (Papilionoidea), Neuropterida, Opiliones, Plecoptera and Trichoptera. The database contains 1 006 768 records with associated geographic data (at 10 × 10 km grid squares), compiled from various sources and validated by specialists in each group. These records cover 4612 species and subspecies, spanning from the 18th century to February 2024. This database represents an unprecedented advance in understanding the distribution of Iberian arthropod biodiversity and its conservation. The complete dataset is provided in Darwin Core Archive format and is available through the Global Biodiversity Information Facility (GBIF).

Keywords: arthropods; insects; Iberian Peninsula; ibero-balearic; occurrence

IberArthro: Base de datos online gratuita que recopila datos taxonómicos y de distribución de los artrópodos iberoibaleares

Resumen: Los artrópodos constituyen la mayor parte de la biodiversidad global, siendo el filo de metazoos más diverso del planeta. A pesar de la relevancia de los artrópodos, nuestro conocimiento sobre ellos sigue siendo limitado, en parte debido a la escasez de datos faunísticos fiables. La península ibérica forma parte de uno de los puntos calientes de biodiversidad global, la Cuenca Mediterránea. Sin embargo, se sabe muy poco sobre la ecología y la distribución espacial de muchas especies de artrópodos ibéricos. Nuestro trabajo presenta una base de datos sobre la distribución de artrópodos iberoibaleares (la base de datos IberArthro) pertenecientes a 10 grupos taxonómicos: Coleoptera (acuáticos), Coleoptera (Scarabaeoidea), Collembola, Hymenoptera (Apoidea), Lepidoptera (Noctuoidea), Lepidoptera (Papilionoidea), Neuropterida, Opiliones, Plecoptera y Trichoptera. La base de datos contiene 1 006 768 registros con datos geográficos asociados (en cuadrículas de 10 × 10 km), recopilados a partir de diversas fuentes y validados por especialistas de cada grupo.

Estos registros cubren 4612 especies y subespecies, que abarcan desde el siglo XVIII hasta febrero de 2024. Esta base de datos representa un avance sin precedentes en el conocimiento de la distribución de la biodiversidad ibérica de artrópodos y su conservación. El conjunto de datos completo se proporciona en formato Darwin Core Archive y está disponible a través del Global Biodiversity Information Facility (GBIF).

Palabras clave: artrópodos; insectos; península ibérica; ibero-balear; presencia

Background and Extended Abstract

Arthropods constitute the bulk of global biodiversity, being the most diverse metazoan phylum on the planet (Barnes et al. 2009; Wilson 2016). Their ecological importance is crucial in terrestrial ecosystems, as they are involved in a multitude of processes essential for life on Earth, including pollination, nutrient cycling or biological control (Gurr et al. 2003; Gullan and Cranston 2005; Allsopp et al. 2008; Boix et al. 2024). A global study of biodiversity or conservation cannot be approached without considering arthropods. However, this very important fraction of biodiversity has been systematically ignored in conservation programs (Samways 2005).

Despite the relevance of arthropods, our knowledge about them remains limited, partly due to the scarcity of reliable faunistic data, a concept known as the "Wallacean shortfall" (Lomolino 2004; Hortal et al. 2015; García-Roselló et al. 2023). Although such knowledge is growing rapidly, many basic aspects of their biology, ecology, and distribution are still poorly understood (e.g., Yadav 2003; Diniz-Filho et al. 2010; Sánchez-Fernández et al. 2021). To address these knowledge gaps and guide effective conservation actions, it is essential to compile taxonomic and distributional data for the different groups of organisms that comprise the arthropods (Martín-Piera and Lobo 2000).

The Iberian Peninsula is part of one of the global biodiversity hotspots, the Mediterranean Basin (Myers et al. 2000; Pascual et al. 2011). However, very little is known on the ecology and spatial distribution of many Iberian arthropod species. Although there is a relatively large number of primary biodiversity records available through the main dataset aggregators such as GBIF, some of these records may lack adequate taxonomic resolution or georeferencing accuracy (Anderson et al. 2016; Marcer et al. 2021, 2022) as data provenance is varied, with data from vouchers specimens in natural history collections being increasingly superseded by observational data. Mistakes may also arise especially within data coming from public online observation platforms, as many arthropod species cannot be properly identified from photography. Thus, the main aim of this project is to systematize accurate information on the occurrences of Iberian arthropods facilitating therefore future conservation actions.

Ten species records databases of several taxonomic groups of arthropods from the Iberian Peninsula and Balearic Islands were combined into a single, homogeneous, and carefully curated database, IberArthro. This resource is designed to provide a valuable resource for researchers, public administrations, conservationists and wildlife enthusiasts. Thus, our objective is to promote open access to the distributional data of arthropods from the Iberian Peninsula and Balearic Islands (Fig. 1) gathered in the IberArthro database.

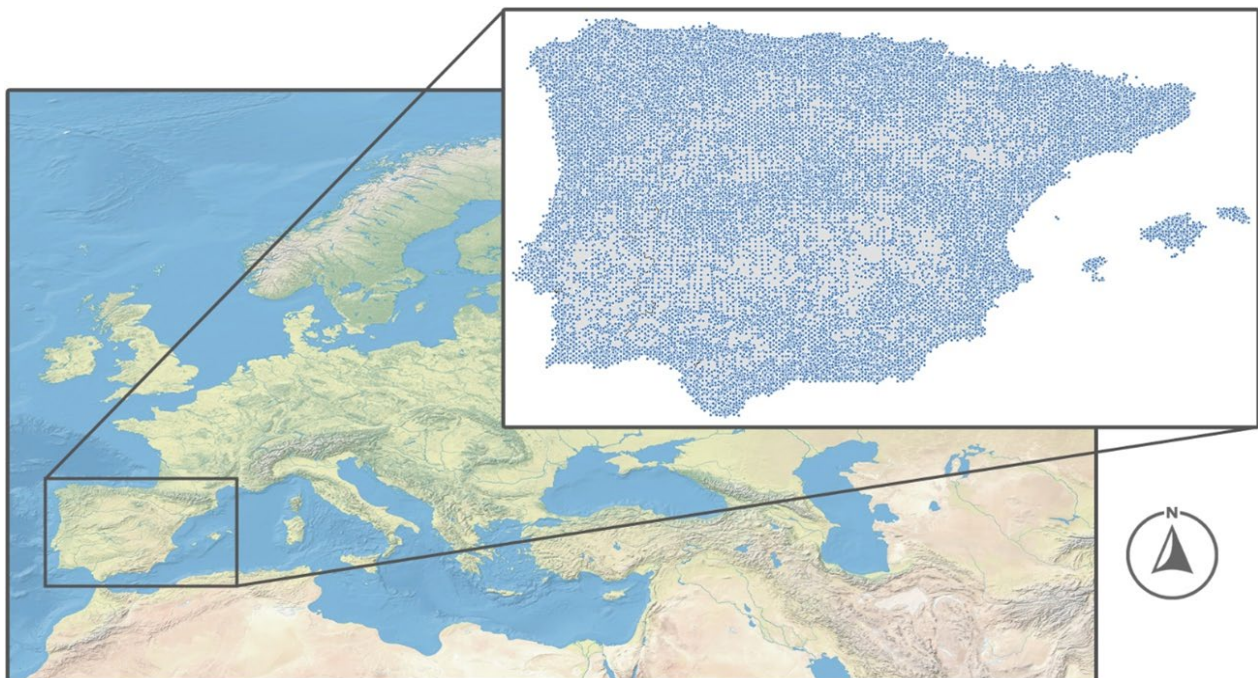


Figure 1. Location of the study area and a detail of the Iberian Peninsula and Balearic Islands. Blue circles represent the location of database records.

Figura 1. Ubicación del área de estudio y detalle de la Península Ibérica y las Islas Baleares. Los círculos azules representan la ubicación de los registros de la base de datos.

The data set is a compilation of 10 taxonomic groups, organized into homogeneous units named and differentiated as follows: Coleoptera (Aquatic), Coleoptera (Scarabaeoidea), Collembola, Hymenoptera (Apoidea), Lepidoptera (Noctuoidea), Lepidoptera (Papilionoidea), Neuropterida, Opiliones, Plecoptera and Trichoptera. This database represents the most comprehensive information available for these arthropods groups in the study area. Each record contains at least information on the species name, locality, geographical coordinates (with a minimum resolution of 10 x 10 km), source, and the 71% of them also containing information on the date.

Older outdated versions of some databases were previously published; including Aquatic Coleoptera (Millán et al. 2014; Sánchez-Fernández et al. 2015); Scarabaeoidea dung beetles (Lobo 2020); Neuropterida (Montserrat and Triviño 2013; Monserrat 2022), Apoidea Hymenoptera (Bartomeus et al. 2022), Papilionoidea Lepidoptera (García-Barros et al. 2004), some of them without access to digital species records.

The database contains 1 006 768 records with associated geographic data for 4612 species belonging to 110 families (Table 1; Fig. 2 and Fig. 3). These records cover most of the Iberian Peninsula, with some differences among taxonomic groups (Fig. 2). The database for Papilionoidea (Lepidoptera) contributed the largest number of records, 465 217 (46.2%), followed by Noctuoidea (Lepidoptera) with 252 688 records (25.1%). The families with the highest number of records are Nymphalidae (196 257) and Noctuidae (189 599), both from Lepidoptera, representing 19.5% and 18.8% of the total records, respectively (Fig. 3). In terms of species richness, the databases of Apoidea (Hymenoptera) and Collembola contain records for 940 and 935 species, respectively.

Table 1. Number of records, species and different families included in the database of each taxonomic group. * Non-Exhaustive database. Do not include more than one record of the same species in the same place.

Tabla 1. Número de registros, especies y familias diferentes incluidas en la base de datos de cada grupo taxonómico. * Base de datos no exhaustiva. No incluye más de un registro de la misma especie en el mismo lugar.

Groups	Records	Species	Order	Families
Coleoptera (Aquatic)	64 400 (6.39 %)	492	Coleoptera	Dryopidae Dytiscidae Elmidae Gyrinidae Haliplidae Helophoridae Hydraenidae Hydrochidae Hydrophilidae Hydroscaphidae Hygrobiidae Noteridae Sphaeriidae
Coleoptera, Scarabaeoidea	62 801 (6.24 %)	223	Coleoptera	Geotrupidae Hybosoridae Ochodaeidae Scarabaeidae
Collembola	28 244 (2.81 %)	935	Collembola	Actaletidae Arrhopalitidae Bourletiellidae Cyphoderidae Dicyrtomidae Entomobryidae Hypogastruridae Isotomidae Katiannidae Neanuridae Neelidae Odontellidae Oncopoduridae Onychiuridae Orcheselidae Paronellidae Poduridae Sminthuridae Sminthurididae Tomoceridae Tullbergiidae
Hymenoptera, Apoidea	75 610 (7.51 %)	940	Hymenoptera	Adrenidae Apidae Colletidae Halictidae Megachilidae Melittidae

Continuation Table 1. / Continuación Tabla1.

Groups	Records	Species	Order	Families
Lepidoptera, Noctuoidea	252 688 (25.09 %)	852	Lepidoptera	Erebidae Euteliidae Noctuidae Nolidae
Lepidoptera, Papilionoidea	465 217 (46.21 %)	241	Lepidoptera	Hesperiidae Lycaenidae Nymphalidae Papilionidae Pieridae Riodinidae
Neuropterida*	12 237 (1.22 %)	205	Megaloptera	Sialidae
			Neuroptera	Ascalaphidae Berothidae Chrysopidae Coniopterygidae Crocidae Dilaridae Hemerobiidae Mantispidae Myrmeleontidae Nemopteridae Nevrorthidae Osmyliidae Sisyridae
			Raphidioptera	Inocelliidae Raphidiidae
Opiliones	9415 (0.94 %)	137	Opiliones	Buemarinoidea Cladonchiidae Dicranolasmatidae Ischyropsalididae Nemastomatidae Phalangiidae Sabaconidae Sclerosomatidae Sironidae Trogulidae
Plecoptera	13 481 (1.34 %)	162	Plecoptera	Capniidae Chloroperlidae Leuctridae Nemouridae Perlidae Perlodidae Taeniopterygidae
Trichoptera	22 675 (2.25 %)	425	Trichoptera	Apataniidae Beraeidae Brachycentridae Calamoceratidae Ecnomidae Glossosomatidae Goeridae Helicopsychidae Hydropsychidae Hydroptilidae Lepidostomatidae Leptoceridae Limnephilidae Odontoceridae Philopotamidae Phryganeidae Polycentropodidae Psychomyiidae Ptilocolepidae Rhyacophilidae Sericoxematidae Thremmatidae
Total	1 006 768 (100 %)	4612	10	110

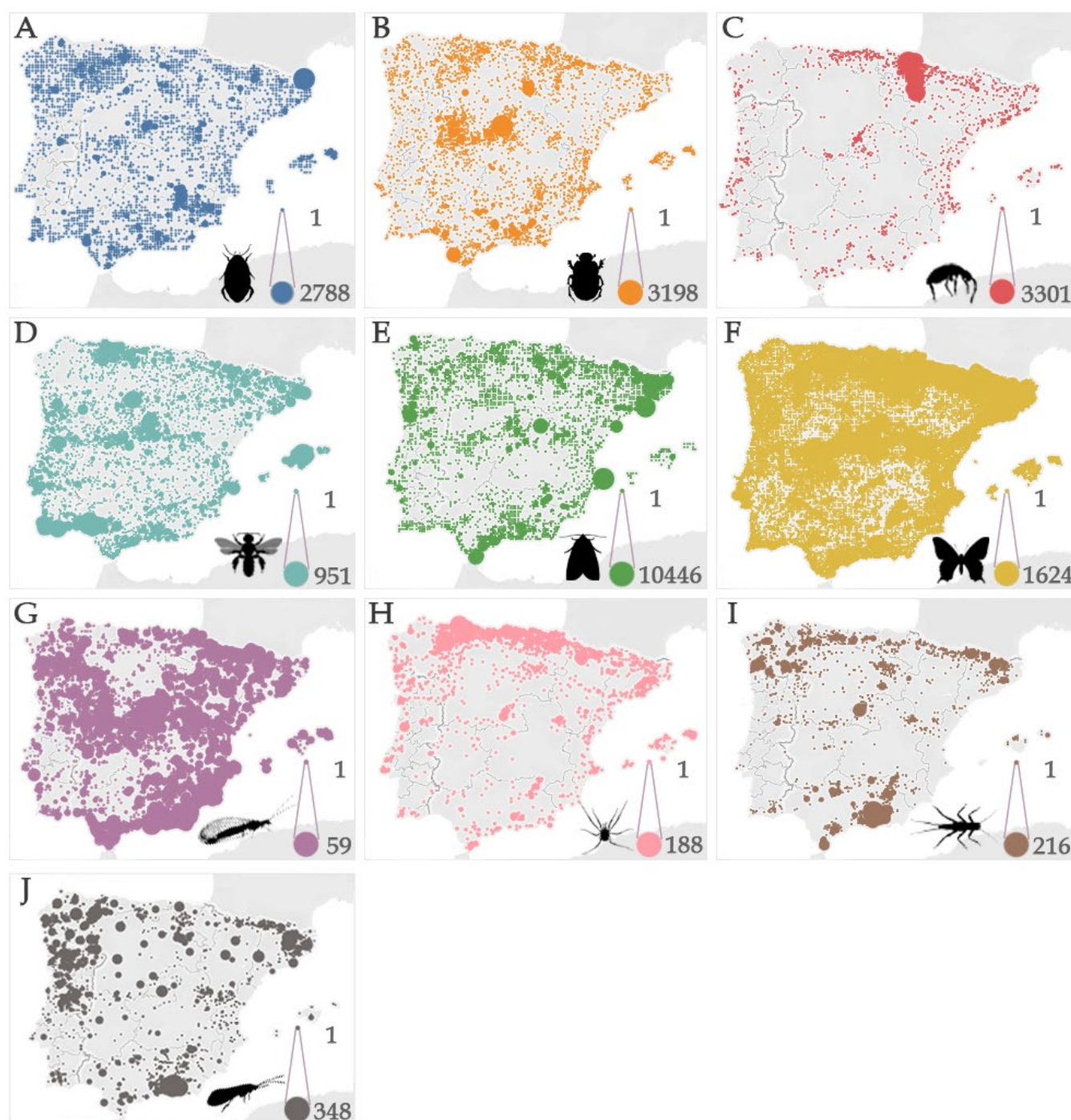


Figure 2. Distribution of the number of records for each taxonomic group. A: Coleoptera (water beetles); B: Coleoptera.Scarabaeoidea; C: Collembola; D: Hymenoptera, Apoidea; E: Lepidoptera, Noctuoidea; F: Lepidoptera, Papilionoidea; G: Neuropterida; H: Opiliones; I: Plecoptera; J: Trichoptera.

Figura 2. Distribución del número de registros para cada grupo taxonómico. A: Coleoptera (coleópteros acuáticos); B: Coleoptera.Scarabaeoidea; C: Collembola; D: Hymenoptera, Apoidea; E: Lepidoptera, Noctuoidea; F: Lepidoptera (Papilionoidea); G: Neuropterida; H: Opiliones; I: Plecoptera; J: Trichoptera.

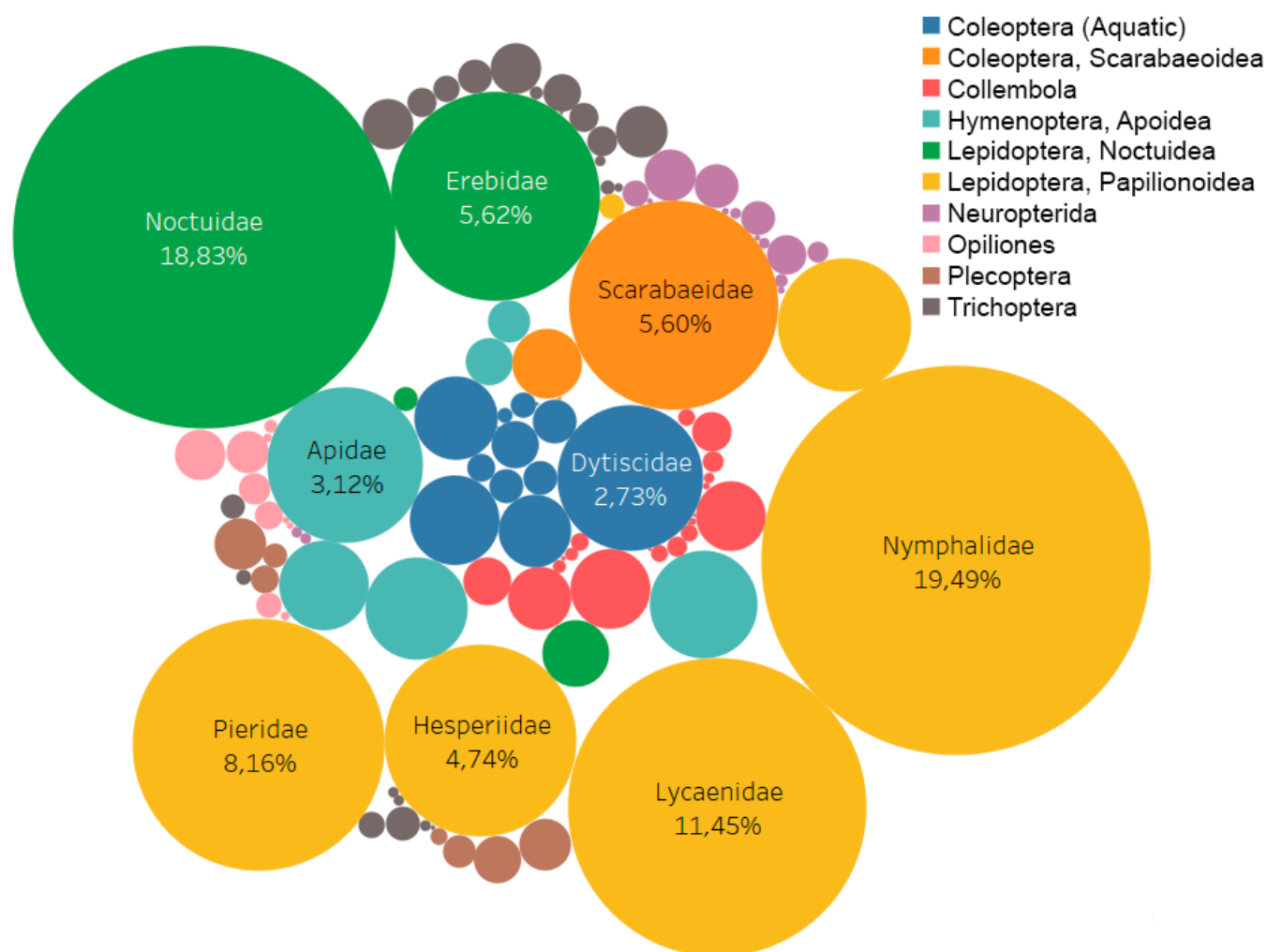


Figure 3. Most recorded families in the IberArthro database. Circle size is proportional to the number of database records for each family.

Figura 3. La mayoría de las familias registradas en la base de datos de IberArthro. El tamaño del círculo es proporcional al número de registros de la base de datos para cada familia.

We are aware that a database of this scale, compiled from various sources and methodologies, may contain a certain number of errors, which is expected in cases of this nature. Errors may arise from incorrect species identification - though this has been minimized by relying on expert-validated databases - or from inaccurate georeferencing of localities, which has also been mitigated through a review process of individual maps. In any case, the conducted sampling effort presents clear biases, which are related to idiosyncratic factors associated with the place of residence or work of the experts specializing in each taxonomic group (Sánchez-Fernández et al. 2022).

This database could support a wide range of studies in disciplines such as ecology, biogeography, and conservation biology, including: i) the identification of current distribution patterns of Iberian arthropods (and future change estimates), ii) the characterization of chorotypes and common biogeographical regions for arthropods, iii) the detection of biodiversity hotspots, as well as the estimated effects of climate change or the assessment of the effectiveness of protected areas in conserving these species. Additionally, it may serve as a foundation for applied studies, such as identifying areas where specific impacts or damages may occur due to the presence of certain species (e.g., pests, disease transmission, etc.).

Material and methods

Sampling description

The data were sourced from a variety of origins, including literature, museum and private collections, PhD theses, fieldwork data, validated citizen contributions, and other unpublished research sources. Some of the most important museums and collections contributing to this dataset include: National Museum of Natural Sciences (MNCN, CSIC, Madrid), Zoology Museum of the University of Navarra (MZNA), Institute of Evolutionary Biology (IBE, CSIC-UPF, Barcelona), Autonomous University of Barcelona (UAB), University of Granada (UGR), University of Murcia (UM), Autonomous University of Madrid (UAM), and University of Castilla-La Mancha (UCLM). Additionally, approximately 8% of all data were obtained from online citizen contributions platforms, including "iNaturalist" (<https://www.inaturalist.org>), "Biodiversidad Virtual" (<https://www.biodiversidadvirtual.org>),

"Observation" (<https://www.observation.org>) and "Fotografía y Biodiversidad" (<https://www.fotografiaybiodiversidad.org>). Noctuoidea must be excluded from this set, due to the impossibility of contrasting the numerous doubts raised by the identification of a high percentage of records. This dataset was generated within the framework of the following project: "Hacia la conservación de artrópodos ibéricos usando herramientas digitales (Digit_Artro)" (TED2021130328B100) funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/ PRTR.

Step description

Each one of the 10 taxonomic groups was compiled separately, following common criteria but with specific characteristics appropriate to each group.

Step by step:

1. Collect primary data sources.
2. Analyze the quality and geographic integrity of all records, with verification by expert taxonomists for each group.
3. Georeferencing. We work with precise coordinates for those records where possible. For the rest, we assigned them to 10 x 10 km grid cells of the Military Grid Reference System (MGRS) based on information from the original sources. Records that could be not unambiguously georeferenced were discarded.
4. Validate the data.
5. Elaborate distribution maps.
6. Checking doubtful records.
7. Correct records (taxonomy or spatial coordinates).
8. Create revised distribution maps.
9. Continue digitization by collecting new data for subsequent updated versions.

Afterward, we worked with all the taxonomic groups to ensure the data were standardized into a common and unified Darwin Core format for submission to GBIF. Information about the columns that make up the dataset can be seen in [Table 2](#).

Table 2. Name and description of the columns in the core dataset (occurrence.txt file). This table is supplemented with three fields from the "extendedmeasurementorfact.txt" file, which provides information on individual counts by sex (indicated with an asterisk).

Tabla 2. Nombre y descripción de las columnas del dataset principal (archivo occurrence.txt). Esta tabla está complementada con tres campos que provienen del archivo "extendedmeasurementorfact.txt", donde se proporciona información sobre los conteos de individuos por sexo (indicado con un asterisco).

Name	Description
Id	Identifier of the record, unique for the dataset.
type	The nature or genre of the resource (Dataset).
modified	The most recent date-time on which the resource was changed.
language	A language of the resource (es = Spanish).
license	A legal document giving official permission to do something with the resource (CC-BY-NC 4.0).
datasetID	A digital identifier for the set of data (doi:10.15470/pqq9oc).
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.
collectionCode	The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived (IberArthro).
datasetName	The name identifying the data set from which the record was derived.
ownerInstitutionCode	The name (or acronym) in use by the institution having ownership of the object(s) or information referred to in the record.
basisOfRecord	The specific nature of the data record (human observation, preserved specimen, occurrence, material citation, machine observation).
dynamicProperties	Original collection of provenance.
occurrenceID	An identifier for the occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.
catalogNumber	An identifier (preferably unique) for the record within the data set or collection.
recordNumber	An identifier given to the occurrence at the time it was recorded. Often serves as a link between field notes and an occurrence record, such as a specimen collector's number.
recordedBy	A list (concatenated and separated) of names of people, groups, or organizations responsible for recording the original Occurrence. The primary collector or observer, especially one who applies a personal identifier (recordNumber), should be listed first.
individualCount	The number of individuals present at the time of the occurrence.
lifeStage	The age class or life stage of the organism(s) at the time the occurrence was recorded (larva, pupa, imago, adult).

Continuation Table 2. / Continuación Tabla2.

Name	Description
occurrenceStatus	A list (concatenated and separated) of identifiers of other occurrence records and their associations to this occurrence.
associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the occurrence.
occurrenceRemarks	Comments or notes about the occurrence.
materialSampleID	An identifier for the material sample (as opposed to a particular digital record of the material sample).
eventDate	The date-time or interval during which an event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context.
Year	The four-digit year in which the event occurred, according to the Common Era Calendar. (From 1784 to 2024).
Month	The integer month in which the event occurred.
Day	The integer day of the month on which the event occurred.
verbatimEventDate	The verbatim original representation of the date and time information.
identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned taxon to the subject.
habitat	A category or description of the habitat in which the event occurred.
continent	The name of the continent in which the dcterms:Location occurs.
country	The name of the country or major administrative unit in which the Location occurs.
countryCode	The standard code for the country in which the Location occurs.
stateProvince, county, municipality	Name of the smaller administrative units than country in which the location occurs.
locality	The specific description of the place.
Mínimum/ maximumElevationInMeters	
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system) of the geographic center of a location. Positive values are north of the Equator; negative values are south of it. Legal values lie between -90 and 90, inclusive.
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system) of the geographic center of a location. Positive values are east of the Greenwich Meridian; negative values are west of it. Legal values lie between -180 and 180, inclusive.
geodeticDatum	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in dwc:decimalLatitude and dwc:decimalLongitude are based.
coordinateUncertaintyInMetres	The horizontal distance (in meters) from the given decimal latitude and decimal longitude describing the smallest circle containing the whole of the location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term.
verbatimCoordinates, verbatimCoordinatesSystem, verbatimSRS	The verbatim original spatial coordinates of the location, Coordinats System and SRS.
verbatimIdentification	A string representing the taxonomic identification as it appeared in the original record.
identificationQualifier	A brief phrase or a standard term ("sp.", "cf.", "aff.") to express the determiner's doubts about the identification.
identifiedBy	A list (concatenated and separated) of names of people, groups, or organizations who assigned taxon to the subject.
scientificName	The full scientific name, with authorship and date information if known.
kingdom	The full scientific name of the kingdom in which the taxon is classified.
phylum	The full scientific name of the phylum or division in which the taxon is classified.
class	The full scientific name of the class in which the taxon is classified.
order	The full scientific name of the order in which the taxon is classified.
family	The full scientific name of the family in which the taxon is classified.
taxonRank	The taxonomic rank of the most specific name in the scientific name.
scientificNameAuthorship	The authorship information for the scientific name formatted according to the conventions of the applicable nomenclatural code.
taxonRemarks	Comments or notes about taxon or name.
measurementType*	maleCount o femaleCount.
measurementValue*	Number of individuals in the record of each type (male or female).
measurementUnit*	Individual Count.

Geographic coverage

Description: The Iberian Peninsula and Balearic Islands are located in the southwest of Europe (Fig. 1). This region includes a variety of biomes, climates and landscapes, with altitudes ranging from sea level to 3483 m a.s.l. in the Sierra Nevada (SE Iberia). The project includes all available distributional data from the Iberian Peninsula, including Andorra, Gibraltar (UK), peninsular Spain and Portugal, and the Balearic Islands (Spain).

Coordinates: 35°23'60"N and 43°58'48"N Latitude; 10°2'24"W and 4°48'36"E Longitude.

Taxonomic coverage

Kingdom: Animalia. Phylum: Arthropoda. Class: Hexapoda and Arachnida

Orders: Coleoptera, Collembola, Hymenoptera, Lepidoptera, Megaloptera, Neuroptera, Raphidioptera, Opiliones, Plecoptera and Trichoptera.

Temporal coverage

Data range from 1784-08-01 to 2024-02-25. The earliest records of arthropod occurrences with geolocation information date back to the late 1700s (De Asso 1784) and correspond to some species of butterflies and beetles. It was not until 1840 that records of other orders, such as Hymenoptera (bees) and aquatic Coleoptera, began to appear. The rest of groups started to have records from 1859 (Plecoptera), 1871 (Noctuoidea, Opiliones, Scarabaeoidea), and 1890 (Collembola and Trichoptera) (Fig. 4). Between 1784 and 1923, the number of annual of arthropod records in the study area ranged from 1 to 500, depending on the taxon. However, between 1924 and 1974, this number increased considerably, ranging between 120 and 4650 records across various taxonomic groups. From 1975 onward, the annual number of records exceeded 6000 in most cases, except in 1998, when fewer than 4000 records were recorded. The year 2022 presented a maximum of almost 40 000 records. It should be noted that, although these databases are kept as up to date as possible, they are subject to some degree of lag due to the vast amount of information handled and the constant increase in knowledge sources (Fig. 4).

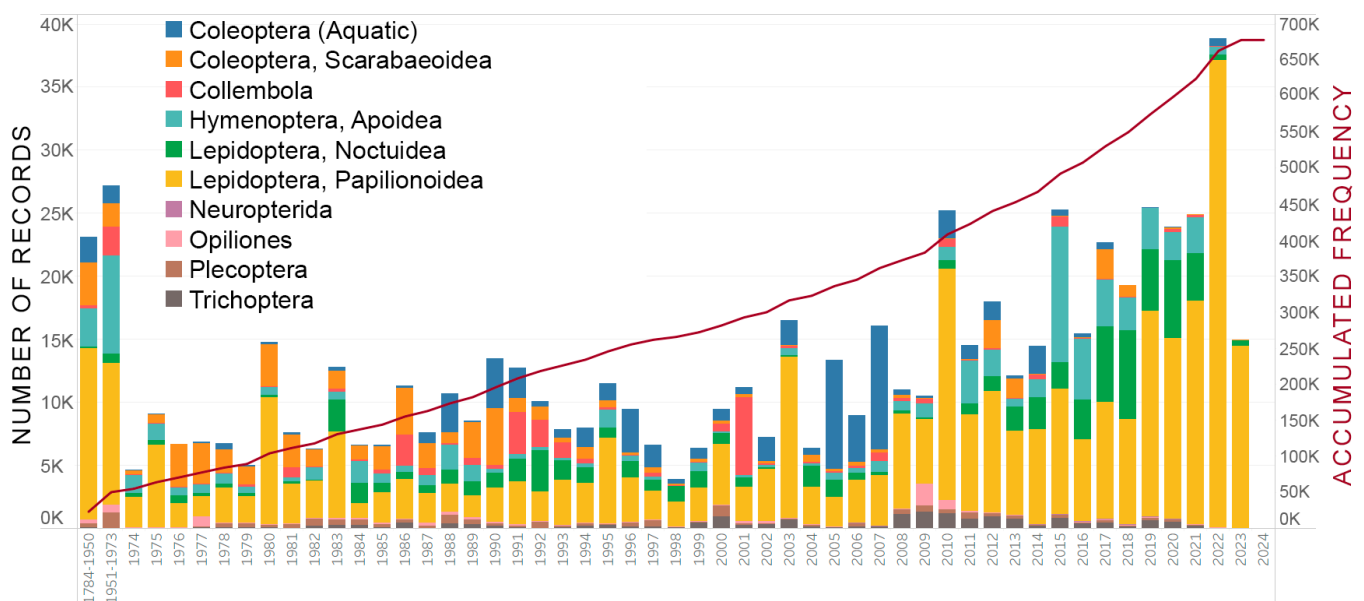


Figure 4. Number of records and temporal tendency of the accumulated number of records in the IberArthro database from each taxonomic group.

Figura 4. Número de registros y tendencia temporal del número acumulado de registros en la base de datos IberArthro de cada grupo taxonómico.

Records and Data Availability and their Use

Description: The database has Darwin Core Archive (DwC-A) format, which includes four files (e.g., occurrences.txt; extendedmeasurementorfact.txt; eml.xml; and metadata.xml). The main file "occurrence.txt" contains 1 006 768 records with associated geographic data (at 10 × 10 km grid squares). These records cover 4612 species and subspecies, spanning from the 18th century to February 2024. This file is complemented by the extendedmeasurementorfact.txt, that allows for adding additional information to records when the standard fields in occurrence.txt are not sufficient, in our case, providing counts of individuals differentiated by sex.

Format: All data were structured according to the Darwin Core Standard for Biodiversity (Wieczorek et al. 2012) and published on the Global Biodiversity Information Facility platform (Mañas-Jordá et al. 2025).

Information about the terms / columns that make up the dataset can be found in **Table 2**.

Usage license: This work is licensed under a Creative Commons Attribution Non-Commercial (CC-BY-NC 4.0) License.

Data package title: IberArthro: A database compiling taxonomic and distributional data on Ibero-Balearic arthropods.

Resource link: <https://doi.org/10.15470/pqq9oc>

Alternative identifiers: <https://www.gbif.org/dataset/73319645-c321-48ed-b923-b3782a203589>

Data set name: IberArthro: A database compiling taxonomic and distributional data on Ibero-Balearic arthropods.

Technical Validation

All the data included in our database have been verified by taxonomists and researchers who are experts in each taxonomic group. Each taxonomic group has an assigned expert or team responsible for conducting a thorough validation of each record. In addition, distribution maps for each species were generated and reviewed multiple times to ensure their accuracy. The current version uploaded to GBIF (2.4) shows a refined dataset after two rounds of technical validation and data curation. In general terms, we used tools and protocols provided by GBIF for evaluating occurrence data quality (<https://www.gbif.org/es/data-quality-requirements-occurrences>) The types of technical checks performed during the two rounds of validation checks for completeness, correctness and consistency to adhere to the Darwin Core Standard and ensure interoperability. Besides, we used the Darwin Core Archive Validator tool (<http://tools.gbif.org/dwca-validator/>) to check whether the dataset meets Darwin Core specifications.

Author's contribution

SMJ: Wrote the first draft of the paper, built the maps and tables, coordinated the network of experts who provided data to compile the dataset, database design, database construction, data compilation. JLY: Project development, collaborate with specimens' ID, collaborated with database construction, data compilation, provided data of Lepidoptera, Noctuoidea. JML: Project development, collaborate with specimens' ID, collaborated with database construction, data compilation, provided data of Coleoptera, Scarabaeoidea. EGR, AJGM: Project development with computing skills (IT-savvy project development). CP: Collaborated with specimens' ID, Data compilation, provided data of Opiliones. VJM: Collaborated with specimens' ID, Data compilation, provided data of Neuropterida. AA, DG, EB: Collaborated with specimens' ID, Data compilation, provided data of Collembola. IB: Data compilation provided data of Hymenoptera, Apoidea. AM: Collaborated with specimens' ID, Data compilation, provided data of Coleoptera (Aquatic). EGB, HR, MLM: Collaborated with specimens' ID, Data compilation, provided data of Lepidoptera, Papilionoidea. RA, NB, MG, JM, CSC: Collaborated with specimens' ID, Data compilation, provided data of Trichoptera. MJLR, JMTF, CSC: Collaborated with specimens' ID, Data compilation, provided data of Plecoptera. DSF: Project leader, project design, project development and collaborated with the writing and revising of the manuscript, collaborated with specimen's ID, Data compilation, provided data of Coleoptera (Aquatic).

Financing, required permits, potential conflicts of interest and acknowledgments

We are grateful to Katia Cezón (Spanish GBIF node—CSIC) for technical support. We also thank all who provided unpublished records and bibliographic references to compile the dataset. DS-F is funded by a postdoctoral contract from the Spanish “Ministerio de Ciencia e Innovación” (“Ramón y Cajal” program, [RYC2019-027446-I]). This paper has been conducted within the framework of the Digit_Artro project (TED2021130328B100), funded by MCIN/AEI /10.13039/501100011033 and by the European Union NextGenerationEU/ PRTR and supported by the projects Ref. PID2021-124187NB-I00 and 3088153 funded by the Spanish Ministry of Science and Innovation and the Spanish Ministry for Ecological Transition and the Demographic Challenge – NextGeneration EU and by the European Union NextGenerationEU, respectively.

The authors declare no conflict of interest.

References

- Allsopp, M., De Lange, W., Veldtman, R. 2008. Valuing Insect Pollination Services with Cost of Replacement. *PLoS ONE* 3 (9). <https://doi.org/10.1371/journal.pone.0003128>
- Anderson, R., Araújo, M., Guisan, A., Lobo, J., Martínez-Meyer, E., Peterson, A., Soberón, J. 2016. Final report of the task group on GBIF data fitness for use in distribution modelling. *Global Biodiversity Information Facility* 1-27.
- Barnes, R.S.K., Calow, P., Olive, P.J.W., Golding, D.W., Spicer, J.I. 2009. *The Invertebrates A Synthesis*. 3rd ed. John Wiley & Sons, Ltd., Hoboken.
- Bartomeus, I., Lanuza, J., Wood, T., Carvalheiro, L., Molina, F., Collado, M.Á., Aguado-Martín, L.O., et al. 2022. Base de datos de abejas ibéricas. *Ecosistemas* 31(3): 2380 <https://doi.org/10.7818/ECOS.2380>
- Boix, D., Bonada, N., Muñoz, I., Baquero, E., Jordana, R., Cunillera-Montcusí, D., Tornero, I., et al. 2024. Class Hexapoda: general introduction. In: Maasri A, Thorp J (Eds) *Identification and ecology of freshwater arthropods in the Mediterranean Basin*. 225-281 pp. [ISBN 978-0-12-821844-0]. <https://doi.org/10.1016/B978-0-12-821844-0.00019-3>
- De Asso, I. 1784. *Introductio in oryctographiam, et zoologiam Aragoniae. Accedit Enumeratio stirpium in eadem regione noviter detectarum*. Autopublished, Amsterdam.

- Diniz-Filho, J.A.F., De Marco, Jr P., Hawkins, B. 2010. Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography. *Insect Conservation and Diversity* 3(3): 172-179. <https://doi.org/10.1111/j.1752-4598.2010.00091.x>
- García-Barros, E., Munguira, M., Martín Cano, J., Romo Benito, H., García-Pereira, P., Maravalhas, E. 2004. *Atlas de las mariposas diurnas de la Península Ibérica e islas Baleares (Lepidoptera: Papilionoidea & Hesperioidea)*. Vol. 11. Monografías Sociedad Entomológica Aragonesa, Zaragoza, España. [ISBN 84-932807-5-5].
- García-Roselló, E., González-Dacosta, J., Lobo, J. 2023. The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. *Biological Conservation* 283 <https://doi.org/10.1016/j.biocon.2023.110118>
- Gullan, P.J., Cranston, P.S. 2005. *The insects: an outline of entomology*. 3rd Edition. Blackwell Pub, Malden, MA. USA.
- Gurr, G., Wratten, S., Luna, J.M. 2003. Multi-function agricultural biodiversity: pest management and other benefits. *Basic and Applied Ecology* 4(2): 107-116. <https://doi.org/10.1078/1439-1791-00122>
- Hortal, J. De Bello, F., Diniz-Filho, J., Lewinsohn, T., Lobo, J., Ladle, R. 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 46(1): 523-549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Lobo, J.M. 2020. BANDASCA, Base de Datos sobre SCArabaeidae. Museo Nacional de Ciencias Naturales (CSIC). <https://doi.org/10.15468/AYFQIZ>
- Lomolino, M. 2004. Conservation biogeography. In: Lomolino, M., Heaney, L. (Eds.), *Frontiers of Biogeography: new directions in the geography of nature*, pp 293-296. Sinauer Associates [ISBN 978-0-87893-479-9].
- Mañas-Jordá, S., Acosta, R., Ariño, A., Baquero, E., Bartomeus, I., Bonada, N., Galicia, D., et al. 2025. *IberArthro: A database compiling taxonomic and distributional data on Ibero-Balearic arthropods*. GBIF. <https://doi.org/10.15470/pqq9oc>
- Marcet, A., Haston, E., Groom, Q., Ariño, A., Chapman, A., Bakken, T., Braun, P., et al. 2021. Quality issues in georeferencing: From physical collections to digital data repositories for ecological research. *Diversity and Distributions* 27 (3): 564-567. <https://doi.org/10.1111/ddi.13208>
- Marcet, A., Chapman, A., Wiczorek, J., Xavier Picó, F., Uribe, F., Waller J., Ariño, A. 2022. Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. *Ecography* 2022 (9). <https://doi.org/10.1111/ecog.06025>
- Martín-Piera, F., Lobo, J. 2000. Diagnóstico sobre el conocimiento sistemático y biogeográfico de insectos hiperdiversos (Coleoptera, Hymenoptera y Lepidoptera) en España. In: Martín-Piera F, Morrone J, Melic A (Eds.), *Hacia un proyecto CYTED para el inventario y estimación de la diversidad entomológica en Iberoamérica: PRIBES 200*. 1. Pp. 287-308. m3m-Monografías Tercer Milenio, Sociedad Entomológica Aragonesa (SEA), Zaragoza, España. [ISBN 84-922495-1-X].
- Millán, A., Sánchez-Fernández, D., Abellán, P., Picazo, F., Carbonell, J.A., Lobo, J.M., Ribera, I. 2014. *Atlas de los coleópteros acuáticos de España peninsular*. Ministerio de Agricultura, Alimentación y Medio Ambiente, Madrid, España.
- Monserat, V.J., Triviño, V. 2013. *Atlas de los neurópteros de la Península Ibérica e Islas Baleares (Insecta, Neuroptera: Megaloptera, Raphidioptera, Planipennia) / Atlas of the Iberian and Balearic lacewings (Insecta, Neuroptera: Megaloptera, Raphidioptera, Planipennia)*. 13. Sociedad Entomológica Aragonesa (S.E.A.), Zaragoza, España.
- Monserat, V.J. 2022. Los Neuropterida de la Península Ibérica y Baleares. *Boletín de la Sociedad Entomológica Aragonesa (S.E.A.)* 16: 229-30.
- Myers, N., Mittermeier, R., Mittermeier, C., Da Fonseca, G.B., Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403 (6772): 853-858. <https://doi.org/10.1038/35002501>
- Pascual, L., Luigi, M., Alessandra, F., Emilio, B., Luigi, B. 2011. Hotspots of species richness, threat and endemism for terrestrial vertebrates in SW Europe. *Acta Oecologica* 37 (5): 399-412. <https://doi.org/10.1016/j.actao.2011.05.004>
- Samways, M. 2005. *Insect diversity conservation*. Cambridge University Press, Cambridge.
- Sánchez-Fernández, D., Millán, A., Abellán, P., Picazo, F., Carbonell, J.A., Ribera, I. 2015. Atlas of Iberian water beetles (ESACIB database). *ZooKeys* 520: 147-154. <https://doi.org/10.3897/zookeys.520.6048>
- Sánchez-Fernández, D., Fox, R., Dennis, R.H., Lobo, J. 2021. How complete are insect inventories? An assessment of the british butterfly database highlighting the influence of dynamic distribution shifts on sampling completeness. *Biodiversity and Conservation* 30 (3): 889-902. <https://doi.org/10.1007/s10531-021-02122-w>
- Sánchez-Fernández, D., Yela, J. L., Acosta, R., Bonada, N., García-Barros, E., Guisande, C., Heine, J., et al. 2022. Are patterns of sampling effort and completeness of inventories congruent? A test using databases for five insect taxa in the Iberian Peninsula. *Insect Conservation and Diversity* 15(4): 406-415. <https://doi.org/10.1111/icad.12566>
- Wilson, E. 2016. *Half-earth: our planet's fight for life*. Liveright Publishing Corporation, New York London, UK.
- Wiczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., et al. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PloS one* 7(1), e29715. <https://doi.org/10.1371/journal.pone.0029715>
- Yadav, M. 2003. *Biology of insects*. Discovery Pub. House, New Delhi, India.