

# Base de datos de abejas Ibéricas

## Iberian Bees database

Ignasi Bartomeus<sup>1</sup>, Jose B. Lanuza<sup>1</sup>, Thomas Woods<sup>2</sup>, Luisa Carvalheiro<sup>3</sup>, and a lot of coauthors<sup>4</sup>

(1) Departamento de Ecología Integrativa, Estación Biológica de Doñana (EBD-CSIC), Consejo Superior de Investigaciones Científicas, Avda. Américo Vespucio s/n, E-41092 Sevilla, España.

(2)

(3)

(4)

Autor para correspondencia: I. Bartomeus [nacho.bartomeus@gmail.com]

## Resumen

**Base de datos de abejas Ibéricas.** Las abejas (Anthophila) son un grupo extremadamente diverso con más de 1000 especies descritas en la Península Ibérica. Además, son excelentes polinizadores y proveen numerosos servicios ecosistémicos fundamentales para la mayoría de ecosistemas terrestres. Debido a los diversos cambios ambientales inducidos por el ser humano, existen evidencias del declive de algunas de sus poblaciones para ciertas especies. Sin embargo, sabemos muy poco del estado de conservación de la mayoría de especies y para muchas de ellas apenas sabemos cual es su distribución en la Península Ibérica. En este trabajo presentamos un esfuerzo colaborativo para agrupar y curar una base de datos de ocurrencias en la Península Ibérica e Islas Baleares que permita responder preguntas sobre su distribución, preferencia de hábitat, fenología o tendencias históricas. En total esta base de datos consta por ahora de 103464 registros de 925 especies con un 89% con información georeferenciada recolectada entre 1830 y 2022. Además cada registro tiene información asociada como la localidad de muestreo (88%), identificador y colector de la especie (68%), fecha de captura (55%) y planta donde se recolectó (23%). Creemos que esta base de datos es el punto de partida para conocer y conservar mejor la biodiversidad de abejas en la Península Ibérica e Islas Baleares. Se puede acceder en: XXXX.  
max 300w

## Abstract

**Iberian Bees database.** Bees are a diverse group with more than 1000 species described in the Iberian Peninsula. They have recently received special attention due to their important role as pollinators and providers of ecosystem services. In addition, various rapid human-induced environmental changes are leading to the decline of some of its populations. However, we know very little about the conservation status of most species and for many species, we hardly know what their distribution is in the Iberian Peninsula. Here we present a collaborative effort to collate and curate a database of bee occurrences to answer questions about their distribution, habitat preference, phenology, or historical trends. In total we have accumulated 103464 records in the Iberian Peninsula and the Balearic Islands of 925 different species collected between 1830 and 2022. More than 89% of these records have associated information such as the location and date of capture, and 23% have extra data such as the plant species where it was collected. This database is the starting point to better understand and conserve bee biodiversity in the Iberian peninsula. It can be accessed at: XXXX.

# Palabras clave

Biodiversidad; Apoidea; Polinizadores; Fenología

# Keywords

Biodiversity; Apoidea; Pollinators; Phenology

# Antecedentes y resumen ampliado/

TBTranslated

# Background & Summary (700 words max).

Bees (Hymenoptera, Apoidea) are a diverse group of species with more than 20.000 species described worldwide ((**Ascher?**)). The Iberian Peninsula, with its bee-loved Mediterranean climate, is one of the hotspots of bee diversity with more than 1000 species described to date and counting ((**Ortiz?**)). Unfortunately, despite the high diversity of Iberian bees, we know very little of their diversity and distribution. This is paradoxical, as in recent years bees have been highlighted as a keystone group in the ecosystem due to a key function they provide, the pollination of thousands of plants ((**ollerton?**)), including most crop species ((**klein?**)). In addition, various rapid human-induced environmental changes are leading to the decline of some of its populations ((**goulson?**)). However, the response of bee species to global change is heterogeneous both in space and taxonomically. That is, while some areas and species are declining drastically ((**Burkle?**)), other areas are well conserved ((**Herrera2020?**)) and while some species are in the brink of extinction ((**Cameron?**)) some species are even thriving ((**Ruso?**)). A better description of bee distribution, habitat preference, phenology, or historical trends in the Iberian peninsula would require information on when and where different species occur.

In the absence of standardized monitoring efforts, one may think that detailed information on species occurrences in space and time does not exist. However, there are different sources of useful information. First, natural history museums hold historical collections of amateur naturalists and researchers who collected bees ((**Bartomeus2018?**)). Collectors are really good at classifying and labeling each collected specimen, which means we can retrieve the information stored along the pinned specimens about when and where those were collected. Second, the Iberian peninsula has a rich tradition of pollinator and pollination researchers ((**Archer?**)), which have been collecting occurrence bee data with different aims. Finally, many occurrence records are being made available through internet portals which centralize data hosted in institutions all around the world (e.g. Gbif) or collected by citizen scientists (INaturalist). The challenge is how to digitalize, access, and curate all this different data sources.

Here we compile an Iberian Bee Database of occurrence records by integrating digitalization efforts by leading natural history museums, with more than 100 individual datasets contributed by researchers, and publicly available information on online repositories. After data cleaning and harmonization, we release the first 103464 records in the Iberian Peninsula and the Balearic Islands. Figure one shows the spatial distribution of records, which as expected is biased to certain areas near main cities (Fig 1). Overall, the database contains information of 925 different species, for which the most commonly collected are depicted in Figure 2. The data has been collected between 1830 y 2022, with an increasing trend of the number of records with time, reflecting a renewed interest for documenting this taxa (Figure 3A). Figure 3 also shows the temporal distribution of records within years (Figure 3B). More than 89% of these records have associated information such as the location and date of capture, and 23% have extra data such as the plant species where it was collected. This Iberian Bee Database (v.1.0) will continue to grow, but it is the starting point to better understand and conserve bee biodiversity. It can be accessed at:

Refs require a bibtext

# Material y métodos

## Recopilación de datos originales

*Museos:* Hemos digitalizado el 10% de los especímenes depositados en el Museo Nacional de Ciencias Natural de Madrid, sobretudo las familias Melitidae y Apidae, para los cuales se han revisado la identificación taxonomicas. THOMAS CAN YOU ADD WHICH MUSEUMS DID YOU DIGITALIZED SPECIMENS, THANKS. En el futuro esta previsto continuar con la digitalización del MNCN y añadir otras colecciones hitóricas.

*Proyectos de investigación:* Hemos contactado los principales investigadores en ecología de la polinización de España y Portugal, así como anunciado en las principales listas de distribución temáticas para recoger datos sobre ocurrencia de abejas en la Península Ibérica. Se recogieron 157 estudios con datos ya digitalizados. Estos han sido limpiados y harmobizados (ver más abajo). En el futuro se seguirán añadiendo nuevas contribuciones, así como se rescatarán datos ya publicados en formatos de difícil extracción como PDF.

*Repositorios de internet:* Usando paquetes programáticos de R ((**bartomeus?**), (**ropensciINat?**), (**ropensciGbif?**)) descargamos todos los datos de Gbif (<https://www.gbif.org>) y INaturalist (<https://www.inaturalist.org>) de ocurrencias de abejas en la península Ibérica a fecha 01-03-2021. Estos datos pueden ser actualizados en nuevas versiones automáticamente.

## Tratamiento de datos

Todo el tratamiento de datos se ha hecho de forma reproducible en R (R Core Team, 2021). Primero hemos seleccionado los campos comunes que reflejan que especie y de que sexo, cuando fue colectada, y donde, así como datos de quien fue el colector, quien determino la identidad de la especie, y referencias de donde ha sido publicada. La Tabla 1 ofrece un resumen de los metadatos creados con (**dataspice?**), y que también están disponibles en formato json (indexado por Google datasets) y EML. Debido a la naturaleza variada de este tipo de datos el proceso de limpieza de datos fue importante. De los 108140 datos iniciales, 4676 fueron descartados. Todo el proceso así como los datos descartados pueden consultarse aquí: <https://github.com/ibartomeus/IberianBees>

Primero se reviso toda la taxonomía de las especies cotejándolos con la lista de especies actualizada para la península ibérica ((**Wood?**), (**Ortiz?**)). Solo se han aceptado especies correctamente identificadas a nivel de especie. Se han actualizado más de 0 registros, actualizando todos los sinónimos y corrigiendo errores ortográficos. Segundo se ha recogido País, Provincia y Localidad, así como estandarizado las geolocalizaciones a latitud y longitud en grados decimales (WGS84) cuando estaba disponible así como la precisión de las coordenadas. Tercero, se ha recogido el día mes y año de colección. En caso de no figurar una fecha precisa, se incluye el intervalo. Para los datos de “donde” y “cuando” se han corregido o eliminado datos imposibles (e.g. geolocalizaciones en el mar, especímenes recogidos en el mes 18). Cuarto, se recoge el nombre del colector y determinador del espécimen, así como el número de machos, hembras y obreras cuando se especifica. Finalmente hemos añadido referencias a publicaciones donde se usan estos datos, y otros datos de interés como la planta que estaban visitando, la identificación única de la base de datos local (si la hay), los autores que han proporcionado los datos y cualquier otra nota de interés.

La base de datos resultante se ha asignado a cada entrada un identificador único, y se guarda tanto el nombre de la especie original tal y como fue inicialmente reportada, como el nombre aceptado final después de su limpieza.

## Flujo de trabajo

Toda la base de datos se puede reconstruir de las fuentes originales usando los scripts proporcionados, y por tanto se pueden trazar todas las decisiones tomadas. Se ha creado una plantilla para contribuir con más datos.

Utilizamos R (R Core Team, 2021) y Rmarkdown (Xie, 2014, 2015, 2020; Xie et al., 2018, 2020; Allaire et al., 2021) para todos nuestros análisis.

## Registro y disponibilidad de datos

Todo el proceso de trabajo esta en abierto en Github y los todos los archivos de codigo creados tienen licencia MIT (<https://opensource.org/licenses/MIT>). El uso de datos tiene licencia CC-BY 4.0 (<https://creativecommons.org/licenses/by/4.0/>). La versión depositada junto al envío de este paper corresponde a la versión 1.0., y tanto los datos como los scripts tienen una versión estable depositada en Zenodo y citable usando este DOI. Se usará un sistema de versionado incremental, donde pequeñas correcciones de errores, o incrementos no substanciales de datos recibirán actualizaciones en el primer dígito (v1.1, v1.2, etc...) y actualizaciones mayores en el segundo número (v2.0, v3.0, etc...). Todas las actualizaciones mayores serán depositadas en Zenodo bajo el DOI general:

## Contribución de los autores

IB and JL cleaned and prepared the database. MAC gathered data from the literature. PA digitalized MNCN data, TW and CM provided taxonomic expertise. IB, LGC, TW and JL wrote the paper. All authors contributed data.

## Agradecimientos

Thanks to the thousands of collectors, naturalists, and scientists who contributed to collect, curate and identify this dataset. We thank project EUCLIPO [Luisa, Fill in details] and SAFEGUARD (ref. 101003476 H2020-SFS-2019-2)

### REFERENCIAS

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H. et al. 2021. *rmarkdown: Dynamic Documents for R*.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Xie, Y. 2015. *Dynamic Documents with R and knitr*. 2nd ed. Chapman; Hall/CRC, Boca Raton, Florida.
- Xie, Y. 2014. knitr: A Comprehensive Tool for Reproducible Research in R. En Stodden, V., Leisch, F., Peng, R. D. (eds.), *Implementing Reproducible Computational Research*, Chapman; Hall/CRC.
- Xie, Y. 2020. *knitr: A General-Purpose Package for Dynamic Report Generation in R*.
- Xie, Y., Allaire, J.J., Grolemund, G. 2018. *R Markdown: The Definitive Guide*. Chapman; Hall/CRC, Boca Raton, Florida.
- Xie, Y., Dervieux, C., Riederer, E. 2020. *R Markdown Cookbook*. Chapman; Hall/CRC, Boca Raton, Florida.

TABLA 1

**Tabla 1.** Metadatos explicando el significado y las unidades de cada variable en el set de datos. [Cada tabla debe aportar su correspondiente encabezamiento explicativo. En los Artículos de investigación, de revisión y Comunicaciones breves se aportarán los encabezamientos tanto en castellano como en inglés, en letra Arial 10 y en página independiente. Es importante que sean simples y que no superen el ancho una página DIN A4 vertical. Los originales se deben aportar en formato tabla y no en formato imagen.]

**Table 1.** Metadata explaining for each variable in the data set, its meaning and units.

variableName	description	unitText
Genus	Genus	Categorical
Subgenus	Subgenus	Categorical
Species	Species	Categorical
Subspecies	Subspecies	Categorical
Country	Country of collection	Categorical
Province	Province of collection	Categorical
Locality	Locality as provided by original collectors	Text
Latitude	Latitude	Decimal degrees

variableName	description	unitText
Longitude	Longitude	Decimal degrees
Coordinate.precision	Precision at which the coordinates are provided	Categorical
Year	Year of collection	Year
Month	Month of collection	Month
Day	Day of collection	Day
Start.date	Free text regarding earlier posible date of collection	Text
End.date	Free text regarding later posible date of collection	Text
Collector	Name of the collector	Categorical
Determined.by	Name of the person who identified the specimen	Categorical
Female	Number of females recorded (including queens in social spcies)	number of records
Male	Number of males recorded	number of records
Worker	Number of workers recorded (for social species)	number of records
Not.specified	Number of specimens recorded without sex identification	number of records
Reference.doi	If the data is published, the DOI of the papers where it appears	Text
Flowers.visited	Free text including the flowers that the specimen was visiting	Text
Local_ID	Unique identification in the local collection	id
Authors.to.give.credit	Authors providing the original dataset	Text
Any.other.additional.data	Free text to note any other additional data	Text
Notes.and.queries	Free text to note any other notes or queries	Text
uid	Unique identification in IBD.	id

## PIES DE FIGURA

**Figura 1.** Barplot indicating the 20 more common recorded species in the final data set.

**Figura 2.** Map of the Iberian peninsula indicating where the bee occurrences were sampled.

**Figura 3.** Temporal coverage of (A) collection years and (B) collection months.

## FIGURE LEGENDS

**Figura 1.** Gráfico de barras indicando la abundancia de las 20 especies más abundantes.

**Figura 2.** Mapa de la Península Ibérica donde se muestran las zonas con mayor numero de registros.

**Figura 3.** Cobertura temporal de (A) los años con más ocurrencias, y (B) los meses con más ocurrencias.

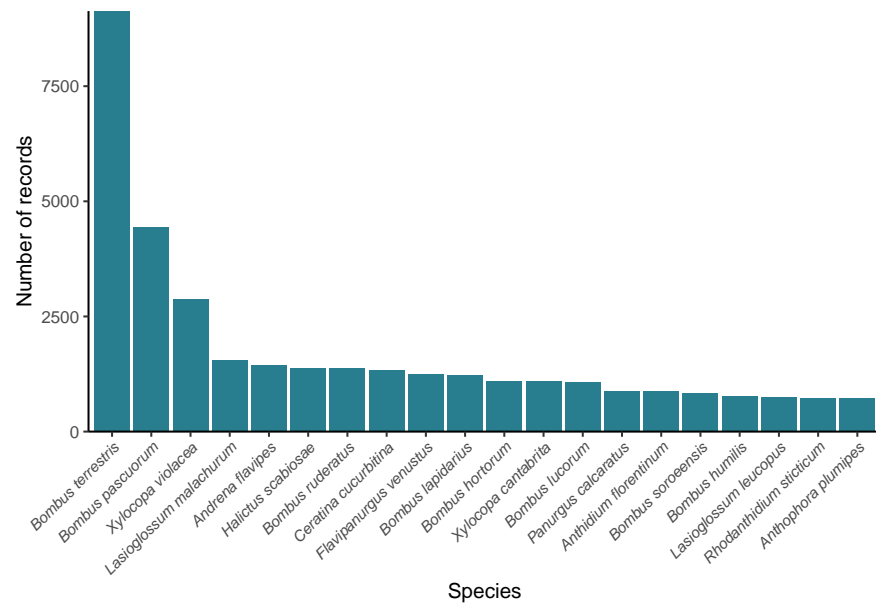


Figure 1: Barplot indicating the 20 more common recorded species in the final data set.

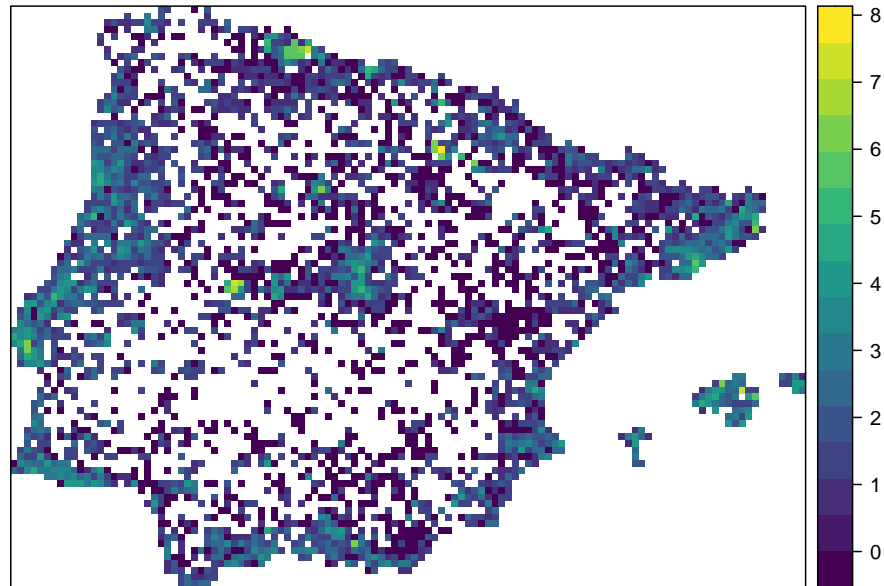


Figure 2: Map of the Iberian peninsula indicating where the bee occurrences were sampled.

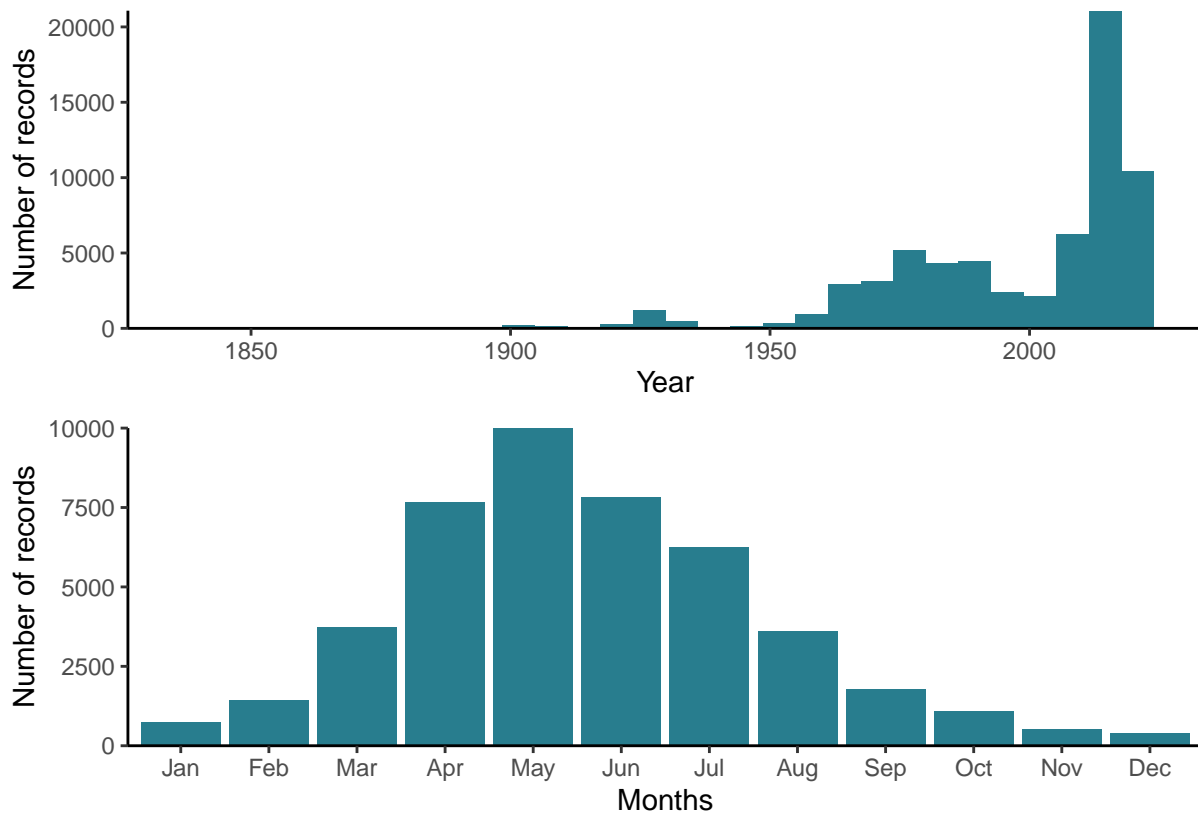


Figure 3: Cobertura temporal de (A) los años con más ocurrencias, y (B) los meses con más ocurrencias.