

# Base de datos de abejas ibéricas

## Iberian bees database

Ignasi Bartomeus<sup>1</sup>  Jose B. Lanuza<sup>1</sup>  Thomas James Wood<sup>2</sup>   
 Luisa Carvalheiro<sup>3,4</sup>  Francisco P. Molina<sup>1</sup>   
 Miguel Ángel Collado Aliaño<sup>5</sup>  Luis Oscar Aguado-Martín<sup>6</sup> David Alomar<sup>7</sup>  
 Marián Álvarez Fidalgo<sup>8</sup>  Piluca Álvarez Fidalgo<sup>9</sup>  Montse Arista<sup>10</sup>  
 Blanca Arroyo-Correa<sup>1</sup>  Josep D. Asís<sup>11</sup>  Celeste Azpiazu<sup>12,13,14</sup>   
 Laura Baños-Picón<sup>11</sup>  Pedro Beja<sup>15</sup>  Mário Boeiro<sup>16</sup>   
 Paulo A.V. Borges<sup>16</sup>  Guillermo González Bornay<sup>17</sup>  Rafael Carvalho<sup>18</sup>   
 Ramón Casimiro-Soriguer<sup>19</sup>  Sílvia Castro<sup>18</sup>  Joana Costa<sup>18,20</sup>   
 Ian Cross<sup>21</sup> Pilar De la Rúa<sup>22</sup>  Luis Miguel de Pablos<sup>23</sup>   
 Víctor de Paz<sup>11</sup>  Joan Díaz-Calafat<sup>24</sup>  Victoria Ferrero<sup>25</sup>   
 Hugo Gaspar<sup>18</sup>  Guillaume Ghisbain<sup>2</sup>  José María Gómez<sup>26</sup>   
 Carmelo Gómez-Martínez<sup>7</sup>  Miguel Ángel González-Estévez<sup>7</sup>  
 Ruben Heleno<sup>18</sup>  José M. Herrera<sup>27</sup>  Jose I. Hormaza<sup>28</sup>   
 Jose M. Iriondo<sup>29</sup>  Michael Kuhlmann<sup>30</sup>  Paola Laiolo<sup>31</sup>   
 Carlos Lara-Romero<sup>29</sup>  Amparo Lázaro<sup>7</sup>  Jesús López-Angulo<sup>32</sup>   
 Francisco A. López-Núñez<sup>18</sup>  João Loureiro<sup>18</sup>  Ainhoa Magrach<sup>33</sup>   
 Vicente Martínez-López<sup>34</sup>  Carlos Martínez-Núñez<sup>35</sup>  Denis Michez<sup>2</sup>   
 Marcos Miñarro<sup>36</sup>  Ana Montero-Castaño<sup>1</sup>  Bruno Moreira<sup>37</sup>   
 Javier Morente-López<sup>29</sup>  Nacho Noval Fonseca<sup>38</sup>   
 Alejandro Núñez Carbajal<sup>36</sup>  José Ramón Obeso<sup>31</sup>  Concepción Ornos<sup>39</sup>   
 Francisco Javier Ortiz-Sánchez<sup>40</sup>  Daniel Pareja Bonilla<sup>10</sup>   
 Sébastien Patiny<sup>2</sup>  Andreia Penado<sup>15,41,42</sup>  Ana Picanço<sup>16</sup>   
 Emilie F. Ploquin<sup>31</sup> Carla Rego<sup>16</sup>  Pedro J. Rey<sup>35</sup>   
 Elisa Ribas-Marquès<sup>43</sup>  Stuart P.M. Roberts<sup>44</sup> Marta Rodríguez<sup>29</sup>  
 Natalia Rosas-Ramos<sup>11</sup>  Ana M. Sánchez<sup>29</sup>  Silvia Santamaría<sup>29</sup>   
 Estefanía Tobajas<sup>11</sup>  José Tormos<sup>11</sup>  Félix Torres<sup>11</sup>   
 Alejandro Trillo<sup>1</sup>  Javier Valverde<sup>1</sup>  Montserrat Vilà<sup>1</sup>   
 Elisa Viñuela<sup>45</sup> 

- (1) Estación Biológica de Doñana (EBD-CSIC), Avenida Américo Vespucio 26, E-41092 Sevilla, Spain.
- (2) Laboratory of Zoology, University of Mons, Avenue du Champs de Mars 6, 7000 Mons, Belgium.
- (3) Centre for Ecology, Evolution and Environmental Changes (cE3c), edificio C2, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal.
- (4) Departamento de Ecologia, Universidade Federal de Goiás, Av. Esperança s/n, 74690-900 Goiânia, Brazil.
- (5) Universidad de Sevilla, Escuela Técnica Superior de Ingeniería Informática, Av. Reina

Mercedes s/n, 41012 Sevilla, Spain.

- (6) Andrena Initiatives and Environmental Studies, Nueva del Saliente 1-Bis, 47328 Valladolid, Spain.
- (7) Instituto Mediterráneo de Estudios Avanzados (IMEDEA; UIB-CSIC), C/ Miquel Marquès 21, 07190 Esporles, Illes Balears, Spain.
- (8) Independent researcher, E-33006 Oviedo, Asturias, Spain.
- (9) Museo Nacional de Ciencias Naturales, C/ de José Gutiérrez Abascal 2, E-28006 Madrid, Spain.
- (10) Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Avenida de la Reina Mercedes 6, 41012 Sevilla, Spain.
- (11) Departamento de Biología Animal (Área de Zoología), Facultad de Biología, Universidad de Salamanca, 37007 Salamanca, Spain.
- (12) Institute of Evolutionary Biology, Universitat Pompeu Fabra, 08034 Barcelona, Spain.
- (13) Centro de Investigación Ecológica y Aplicaciones Forestales, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.
- (14) Universidad Politécnica de Madrid, 28040 Madrid, Spain.
- (15) CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal.
- (16) Centre for Ecology, Evolution and Environmental Changes, Azorean Biodiversity Group, Faculty of Agricultural Sciences and Environment, University of the Azores, PT-9700-042 Angra do Heroísmo, Portugal.
- (17) Centro Universitario de Plasencia, Universidad de Extremadura, Virgen del Puerto 2, 10600 Plasencia, España.
- (18) Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas 3000-456 Coimbra, Portugal.
- (19) Departamento de Biología – Instituto de Investigación Vitivinícola y Agroalimentaria, University of Caádiz, Campus Río San Pedro, ES11510 Puerto Real, Spain.
- (20) Linking Landscape, Environment, Agriculture and Food (LEAF), Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisboa, Portugal.
- (21) 16 Briantspuddle, Dorchester, Dorset, United Kingdom.
- (22) Facultad de Veterinaria, Departamento de Zoología y Antropología Física, Universidad de Murcia, 30100 Murcia, Spain.
- (23) Grupo CTS-183 Bioquímica y Parasitología Molecular, Departamento de Parasitología, Universidad de Granada, 18071 Granada, Spain.
- (24) Southern Swedish Forest Research Centre, Swedish University of Agricultural Sciences, 230 53 Alnarp, Sweden.
- (25) Departamento de Biodiversidad y Gestión Ambiental, Facultad de Ciencias Biológicas y Ambientales, Universidad de León, Campus de Vegazana s/n, 24071 León, Spain.
- (26) Departamento de Ecología Funcional y Evolutiva, Estación Experimental de Zonas Áridas (EEZA-CSIC), Ctra. de Sacramento s/n, La Cañada de San Urbano, 04120 Almería, Spain
- (27) Instituto Mediterráneo para a Agricultura, Ambiente e Desenvolvimento, Universidade de Évora, 7006-554 Évora, Portugal.

- (28) Instituto de Hortofruticultura Subtropical y Mediterránea la Mayora (IHSM la Mayora - CSIC - UMA), Avenida Dr. Wienberg, s/n, 29750 Algarrobo-Costa, Málaga, Spain.
- (29) Universidad Rey Juan Carlos, C/ Tulipan s/n, E-28933 Móstoles, Madrid, Spain.
- (30) Zoological Museum, University of Kiel, Hegewischstr. 3, D-24105 Kiel, Germany.
- (31) Instituto Mixto de Investigación en Biodiversidad, Universidad de Oviedo, Campus de Mieres, E-33006 Mieres, Asturias, Spain.
- (32) Department of Environmental Systems Science, Swiss Federal Institute of Technology Zurich (ETH), 8092 Zurich, Switzerland.
- (33) Basque Centre for Climate Change, Edif. Sede 1, 1º, Parque Científico UPV-EHU, Barrio Sarriena s/n, 48940 Leioa, Spain.
- (34) Institute of Infection, Veterinary and Ecological Sciences, Department of Evolution, Ecology and Behaviour, University of Liverpool, L69 7ZB Liverpool, United Kingdom.
- (35) Departamento de Biología Animal, Biología Vegetal y Ecología, Universidad de Jaén, E-23071 Jaén, Spain.
- (36) Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Ctra AS-267, PK 19, E-33300 Villaviciosa, Asturias, Spain.
- (37) Centro de investigaciones sobre desertificación (CIDE-CSIC/UV/GV), Carretera Moncada-Náquera km 4,5 s/n 46113 Moncada, Valencia, Spain.
- (38) Independent researcher, E-33510 Pola de Siero, Asturias, Spain.
- (39) Departamento de Biodiversidad, Ecología y Evolución, Universidad Complutense de Madrid, C/ José Antonio Novais 12, 28040 Madrid, Spain.
- (40) Grupo de Investigación “Transferencia de I+D en el Área de Recursos Naturales”, Universidad de Almería, E-04120 La Cañada de San Urbano, Almería, Spain.
- (41) School of Life Sciences, University of Sussex, BN1 9QG, Brighton, United Kingdom.
- (42) BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal.
- (43) Independent researcher, Palma, Mallorca, Illes Balears, Spain.
- (44) Agroecology Lab, Université libre de Bruxelles, Campus de la Plaine, B-1050 Brussels, Belgium.
- (45) Unidad de Protección de Cultivos, Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

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

## Resumen

**Base de datos de abejas ibéricas.** Las abejas son un grupo extremadamente diverso con más de 1,000 especies descritas en la península ibérica. Además, son excelentes polinizadores y aportan 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, conocemos muy poco del estado de conservación de la mayoría de especies y de muchas de ellas ignoramos cuál es su distribución en la península ibérica. En este trabajo presentamos un esfuerzo colaborativo para crear una base de datos de ocurrencias de abejas que abarca la península ibérica e islas Baleares que permitirá

resolver cuestiones como la distribución de las diferentes especies, preferencia de hábitat, fenología o tendencias históricas. En su versión actual, esta base de datos contiene un total de 87,684 registros de 923 especies recolectados entre 1830 y 2022, de los cuales un 87% presentan información georreferenciada. Para cada registro se incluye información relativa a la localidad de muestreo (89%), identificador y colector de la especie (64%), fecha de captura (54%) y planta donde se recolectó (20%). 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 a estos datos a través del siguiente enlace permanente: <https://doi.org/10.5281/zenodo.6354502>

**Palabras clave:** biodiversidad; Apoidea; Anthophila; polinizadores; fenología; península ibérica; base de datos

## Abstract

**Iberian bees database.** Bees are a diverse group with more than 1,000 species known from the Iberian Peninsula. They have increasingly 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 their true distributions across the Iberian Peninsula. Here we present a collaborative effort to collate and curate a database of Iberian bee occurrences to answer questions about their distribution, habitat preference, phenology, or historical trends. In total we have accumulated 87,684 records from the Iberian Peninsula and the Balearic Islands of 923 different species with 87% of georeferenced records collected between 1830 and 2022. In addition, each record has associated information such as the sampling location (89%), collector and person who identified the species (64%), date of the capture (54%) and plant species where the bees were captured (20%). We believe that this database is the starting point to better understand and conserve bee biodiversity in the Iberian Peninsula. It can be accessed at: <https://doi.org/10.5281/zenodo.6354502>.

**Keywords:** Biodiversity; Apoidea; Anthophila; Pollinators; Phenology

## Background & Summary

Bees (Hymenoptera, Apoidea, Anthophila) are a diverse group of insects with more than 20,000 species described worldwide (Ascher and Pickering, 2020). The Iberian Peninsula, with its bee-loved Mediterranean climate, is one of the hotspots of bee biodiversity with more than 1,000 species recorded to date and counting (Ascher and Pickering, 2020; Ortiz-Sánchez, 2020). Unfortunately, despite the high biodiversity of Iberian bees, we know very little about the species themselves and their distributions. This is paradoxical, as in recent years bees have been highlighted as a keystone group in the ecosystem due to a essential function they provide, the pollination of thousands of plants (Ollerton et al., 2011), including most crop species (Klein et al., 2007). In addition, various rapid human-induced environmental changes are leading to the decline of some bee populations (Goulson et al., 2015). However, the response of bee species to global change is heterogeneous both in space and also taxonomically. That is, while in some areas bee species are declining drastically (Burkle et al., 2013), in other areas are well conserved (Herrera, 2019) and while some species are in the brink of extinction (Cameron et al., 2011) other species are even thriving (Russo, 2016). A better description of bee distributions, habitat preferences, phenologies, 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 extremely useful information. First, natural history museums hold historical collections of amateur naturalists and researchers who collected bees (Bartomeus et al., 2019). Collectors are very 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 et al., 2014), 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. Global Biodiversity Information Facility, GBIF) or collected by citizen scientists (e.g. iNaturalist). The challenge is how to digitalize, access, and curate all these different data sources. Thus, despite the obvious challenges to digitalize, access, and curate all these different data sources, they provide an essential source of information for biodiversity research (Nelson y Ellis, 2019; Hedrick et al., 2020)

Here we have compiled an Iberian bee database of occurrence records. For this, we have digitalize natural history records from museums, compiled more than 100 datasets contributed by individual researchers, and gathered publicly available information from online repositories (see **Fig. 1** for workflow). After data cleaning and harmonization, we release the first 87,684 records from the Iberian Peninsula and the Balearic Islands. **Figure 2** shows the spatial distribution of records, which as expected it is biased to the main urban areas. Overall, the database contains information of 923 different species and in **Figure 3** we show the 20 most frequently collected species. The data has been collected between 1830 and 2022, with an increasing trend of the number of records with time, reflecting a renewed interest for documenting bee species (**Fig. 4A**). As expected, most records were recorded in spring with May being the month with highest number of records (**Fig. 4B**). More than 87% of these records have associated information such as the location and date of capture, and 20% have extra data such as the plant species where bee specimens were collected. Here, we release the version 1.0 of the Iberian bee database with name “IberianBees”. Although this database is a work in progress, we believe that this is a good starting point to better understand and conserve bee biodiversity. It can be accessed at the following permanent link: <https://doi.org/10.5281/zenodo.6354502>.

## Antecedentes y resumen ampliado

Las abejas (Hymenoptera, Apoidea, Anthophila) son un grupo muy diverso con más de 20,000 especies descritas en el mundo (Ascher y Pickering, 2020). Las condiciones climáticas de la península ibérica hacen a esta región geográfica uno de los puntos calientes de biodiversidad de abejas a nivel mundial con más de 1,000 especies descritas hasta la fecha (Ortiz-Sánchez, 2020). A pesar de la gran riqueza de especies de abejas ibéricas, conocemos muy poco de su ecología y distribución. Paradójicamente las abejas son un grupo clave que desarrolla una función vital para la mayoría de ecosistemas terrestres, la polinización de miles de plantas (Ollerton et al., 2011), incluyendo la mayoría de cultivos (Klein et al., 2007). Además los cambios ambientales originados por el impacto del ser humano están ocasionando el declive de algunas de sus poblaciones (Goulson et al., 2015). No obstante, la respuesta de las abejas al cambio global es heterogénea tanto espacial como taxonómicamente. Es decir, existen áreas que están sufriendo un decaimiento poblacional drástico (Burkle et al., 2013), mientras que otras se mantienen bien conservadas (Herrera, 2019) y aunque algunas especies se encuentran al borde de la extinción (Cameron et al., 2011), otras están incluso prosperando (Russo, 2016). Para un mayor conocimiento de la distribución, preferencia de hábitat, fenología, o tendencias históricas de las diferentes especies de abejas es necesaria información detallada del escenario temporal y espacial de aparición de las mismas.

Debido a la ausencia de esfuerzos de muestreo estandarizados se podría pensar que no existe información detallada del registro de especies en el espacio y en el tiempo. Sin embargo, hay diversas fuentes con información muy útil. En primer lugar, los museos de historia natural albergan colecciones históricas de naturalistas amateur e investigadores que capturaron abejas (Bartomeus et al., 2019). Puesto que las colecciones de museos suelen estar muy bien clasificadas y etiquetadas, es posible recopilar información específica sobre la localización y momento de captura de los diferentes especímenes. En segundo lugar, la península ibérica tiene una rica tradición de investigadores sobre polinizadores y ecología de la polinización (Archer et al., 2014) que han recopilado gran cantidad de información de abejas con diferentes objetivos. Finalmente, muchos datos de ocurrencias han sido liberados de forma abierta a través de portales de internet que centralizan datos almacenados por instituciones de todo el mundo (p.ej. GBIF) o datos de ciencia ciudadana (p.ej. iNaturalist). Aunque sigue siendo un desafío digitalizar, acceder y depurar la información procedente de las diferentes fuentes de datos, la información que estas fuentes aportan es esencial para la investigación de la biodiversidad (Nelson y Ellis, 2019; Hedrick et al., 2020)

En este estudio hemos creado una base de datos de abejas ibéricas mediante la unificación de: (i) los esfuerzos de digitalización de museos, (ii) conjuntos de datos de investigadores, e (iii) información de repositorios en línea de libre acceso (ver **Fig. 1** para ver flujo de trabajo). Tras la depuración y homogeneización de los datos, se han liberado los primeros 87,684 registros que abarcan la península ibérica e Islas Baleares. La **Figura 2** muestra la distribución espacial de los registros, observándose, tal y como se esperaba, un sesgo hacia ciertas áreas cercanas a los grandes núcleos urbanos. Esta base de datos contiene información sobre 923 diferentes especies y el número de observaciones de las 20 más comunes se encuentran representadas en la **Figura 3**. Los datos han sido recopilados entre los años 1830 y 2022, con un ascendente número de registros en el tiempo que refleja el creciente interés de documentar a este grupo taxonómico (**Fig. 4A**). Además, como es de esperar, la mayoría de registros tienen lugar en los meses de primavera, siendo mayo el mes que cuenta con un mayor número de registros (**Fig. 4B**). El 87% de estos registros presenta información georreferenciada y el resto de registros al menos contiene información de la localidad del evento de captura u observación del espécimen. También un 54% de los registros aporta la fecha de captura, y un 20% tiene información de la planta donde se capturó u observó el espécimen. Esta es la versión 1.0 de la base de datos de abejas ibéricas con nombre “IberianBees” que, aunque se seguirá actualizando, creemos que es un buen punto de partida para mejorar nuestro conocimiento y contribuir a la conservación de este grupo taxonómico tan importante para nuestros ecosistemas y la producción agrícola. Se puede acceder desde el siguiente link permanente: <https://doi.org/10.5281/zenodo.6354502>.

## Material y métodos

### Recopilación de datos originales

*Museos:* Hemos digitalizado el 10% de los especímenes depositados en el Museo Nacional de Ciencias Naturales de Madrid (MNCN), sobre todo las familias Melittidae y Apidae, para las cuales se han revisado las identificaciones taxonómicas. Además se ha recabado información digitalizada de museos extranjeros incluyendo “Naturalis Biodiversity Center” (Holanda), “Oberösterreichisches Landesmuseum” (Austria) y la “Smithsonian collection” (USA). En el futuro está previsto continuar con la digitalización del MNCN, así como añadir otras colecciones históricas de importancia.

*Proyectos de investigación:* Para recoger datos sobre ocurrencia de abejas en la península ibérica hemos contactado con los principales investigadores en ecología de la polinización de España y Portugal, así como publicitado en las principales listas de distribución temáticas la creación de una base de datos de abejas ibéricas. En total, se recogieron 148 estudios con datos ya digitalizados que han sido depurados y homogeneizados (ver más abajo). En el futuro se seguirán añadiendo nuevas contribuciones y además se rescatarán datos ya publicados en formatos de difícil extracción como PDF.

*Repositorios de internet:* Usando los paquetes de R *rgbif* versión 3.6.0 (Chamberlain et al., 2022) y *rinat* versión 0.1.8 (Barve y Hart, 2022) hemos descargado todas las ocurrencias de abejas de la península ibérica de los repositorios de libre acceso GBIF (<https://www.gbif.org>) e iNaturalist (<https://www.inaturalist.org>) a fecha 14-03-2022. Estos datos pueden ser actualizados en nuevas versiones automáticamente y así incorporar los nuevos registros a la base de datos.

### Tratamiento de datos

Todo el tratamiento de datos se ha hecho de forma reproducible en R (R Core Team, 2021). Primero se han seleccionado los campos comunes que reflejan la especie, el sexo, fecha de captura y localización del registro, además de la referencia de la publicación correspondiente al registro cuando era posible. La **Tabla 1** ofrece un resumen de los metadatos creados con DataSpice (Boettiger et al., 2022), y también están disponibles en formato JSON y XML que cumple el estándar EML -Ecological Metadata Language- (Jones et al., 2019), los cuales están indexados por Google datasets. Debido a la naturaleza variada de este tipo de datos, el proceso de limpieza y homogeneización de datos fue complejo y tedioso. De los 92,865 datos iniciales, 5,181 fueron descartados. Todo el proceso de tratamiento de datos, así como los datos descartados, pueden consultarse tanto en Github (<https://github.com/ibartomeus/IberianBees>) como en Zenodo (<https://doi.org/10.5281/zenodo.6354502>).

En primer lugar se revisó toda la taxonomía de las especies que se cotejó con la lista de especies actualizada para la península ibérica (Ortiz-Sánchez, 2020). Solo se han incluido especies con nombres taxonómicos aceptados en este último listado. En total, se han comprobado más de 489 registros, actualizando todos los sinónimos y corrigiendo errores tipográficos. En segundo lugar, se ha recogido información del país, provincia y localidad en la que se produjo el registro. Además se realizó un proceso de homogeneización de las geolocalizaciones (latitud y longitud) a grados decimales (WGS84) siempre y cuando esta información estuviera presente. Aunque también se incluyó información sobre la precisión de las coordenadas, esta variable estaba raramente disponible para muchas de las coordenadas. En tercer lugar, se ha incluido también información de la fecha de la captura u observación con su respectivo día, mes y año. Para registros sin fechas concretas, se han intentado recuperar un intervalo temporal orientativo. Los datos con fechas y localizaciones erróneas o ilógicas (p.ej. geolocalizaciones en el mar o especímenes recogidos en el mes 18) han sido corregidos o eliminados. En cuarto lugar, se han añadido los nombres de los colectores y personas que identificaron las especies, así como información sobre el número de abejas macho, hembra u obreras cuando se especificaba. Finalmente, se han añadido las referencias correspondientes a cada registro y otros datos de interés como la planta visitada, identificación única de la base de datos local (si la hay), autores que han proporcionado los datos y cualquier otra información relevante si la hubiese.

En la base de datos resultante se ha asignado a cada entrada o registro un identificador único. Además, se ha conservado tanto el nombre de la especie original con el que fue inicialmente identificada, como el nombre aceptado tras la revisión. Sinonimias o nombres de especies que no se encontraron en la lista actualizada de abejas ibéricas fueron consultadas a taxónomos expertos antes de ser descartadas.

## 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. Finalmente, se ha creado una plantilla para contribuir con más datos que puede ser consultada en el repositorio de esta base de datos.

Utilizamos R (R Core Team, 2021) y Rmarkdown (Xie et al., 2018, 2020; Allaire et al., 2020) para todo el procesado, escritura y creación de esta base de datos.

## Registro y disponibilidad de datos

Todos los archivos del código creado tienen licencia MIT (<https://opensource.org/licenses/MIT>) y 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 artículo corresponde a la versión 1.0. de IberianBees y tanto los datos como los scripts, tienen una versión permanente depositada en Zenodo que es citable usando este identificador único: <https://doi.org/10.5281/zenodo.6354503>. Además todo el proceso de trabajo se encuentra disponible en el repositorio de Github: <https://github.com/ibartomeus/IberianBees>. Se usará un sistema de versionado incremental, donde pequeñas correcciones de errores o incrementos no substanciales de datos recibirán actualizaciones en el segundo dígito (v1.1, v1.2, etc.) y actualizaciones con modificaciones considerables tendrán cambios en el primer número (v2.0, v3.0, etc.). Todas estas versiones con grandes cambios serán depositadas en Zenodo bajo el link permanente: <https://doi.org/10.5281/zenodo.6354502>.

## Contribución de los autores

IB, LGC y TW han compilado los datos de museos y propietarios individuales de datos. MAC ayudó con la recogida de datos de la literatura. PA digitalizó los datos de MNCN. TW y CM ayudaron con las revisiones taxonómicas. IB y JBL depuraron y prepararon la base de datos. IB, JBL, LGC y TW escribieron el artículo. Todos los autores aportaron registros a la base de datos.

## Agradecimientos

Gracias a todos los naturalistas y científicos que han contribuido de algún modo en la creación de esta base de datos, este trabajo es el resultado del cariño y esfuerzo de miles de personas. Esta base de datos se ha realizado con la ayuda de los proyectos EUCLIPO (Fundação para a Ciência e a Tecnologia, LISBOA-01-0145-FEDER-028360/EUCLIPO) y SAFEGUARD (ref. 101003476 H2020-SFS-2019-2).

## Referencias

- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H. et al. 2020. *rmarkdown: Dynamic Documents for R*.
- Archer, C.R., Pirk, C.W.W., Carvalheiro, L.G., Nicolson, S.W. 2014. Economic and ecological implications of geographic bias in pollinator ecology in the light of pollinator declines. *Oikos* 123: 401-407.
- Ascher, J., Pickering, J. 2020. Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila).
- Bartomeus, I., Stavert, J., Ward, D., Aguado, O. 2019. Historical collections as a tool for assessing the global pollination crisis. *Philosophical Transactions of the Royal Society B* 374: 20170389.
- Barve, V., Hart, E. 2022. *rinat: Access 'iNaturalist' Data Through APIs*.
- Boettiger, C., Chamberlain, S., Fournier, A., Hondula, K., Krystalli, A., Mecum, B., Salmon, M. et al. 2022. *dataspice: Create Lightweight Schema.org Descriptions of Data*.
- Burkle, L.A., Marlin, J.C., Knight, T.M. 2013. Plant-pollinator interactions over 120 years: loss of species, co-occurrence, and function. *Science* 339: 1611-1615.
- Cameron, S.A., Lozier, J.D., Strange, J.P., Koch, J.B., Cordes, N., Solter, L.F., Griswold, T.L. 2011. Patterns of widespread decline in North American bumble bees. *Proceedings of the National Academy of Sciences* 108: 662-667.
- Chamberlain, S., Barve, V., McGlinn, D., Oldoni, D., Desmet, P., Geffert, L., Ram, K. 2022. *rgbif: Interface to the Global Biodiversity Information Facility API*.
- Goulson, D., Nicholls, E., Botías, C., Rotheray, E.L. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347: 1255-1257.
- Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J. et al. 2020. Digitization and the future of natural history collections. *BioScience* 70: 243-251.
- Herrera, C.M. 2019. Complex long-term dynamics of pollinator abundance in undisturbed Mediterranean montane habitats over two decades. *Ecological Monographs* 89: e01338.
- Jones, M., O'Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T. et al. 2019. Ecological Metadata Language version 2.2.0.
- Klein, A.-M., Vaissiere, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T. 2007. Importance of pollinators in changing landscapes for world crops. *Proceedings of the royal society B: biological sciences* 274: 303-313.
- Nelson, G., Ellis, S. 2019. The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B* 374: 20170391.
- Ollerton, J., Winfree, R., Tarrant, S. 2011. How many flowering plants are pollinated by animals? *Oikos* 120: 321-326.
- Ortiz-Sánchez, F.J. 2020. Checklist de Fauna Ibérica. Serie Anthophila (Insecta: Hymenoptera: Apoidea) en la Península Ibérica e Islas Baleares (edición 2020). *Madrid: Museo Nacional de Ciencias Naturales, CSIC* 2: 83.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Russo, L. 2016. Positive and negative impacts of non-native bee species around the world. *Insects* 7: 69.
- 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.



## **LEYENDA DE TABLA**

**Tabla 1.** Metadatos explicando el significado y las unidades de cada variable en el conjunto de datos.

## **TABLE CAPTION**

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

**Tabla 1**

<b>Variable name</b>	<b>Description</b>	<b>Data type</b>
Genus	Taxonomic rank	Categorical
Subgenus	Taxonomic rank	Categorical
Species	Taxonomic rank	Categorical
Subspecies	Taxonomic rank	Categorical
Accepted.name	Accepted species name	Categorical
Original.name	Original species name	Categorical
Country	Country of collection	Categorical
Province	Province of collection	Categorical
Locality	Locality as provided by original collectors	Text
Latitude	Geographic coordinates	Decimal degrees
Longitude	Geographic coordinates	Decimal degrees
Coordinate.precision	Accuracy of the coordinates provided by the authors	Text
Year	Year of collection	Date (YYYY)
Month	Month of collection	Date (MM)
Day	Day of collection	Date (DD)
Start.date	Approximate range of dates if specific dates are missing	Date (DD-MM-YYYY)
End.date	Approximate range of dates if specific dates are missing	Date (DD-MM-YYYY)
Collector	Name of the collector	Categorical
Determined.by	Name of the person who identified the specimen	Categorical
Female	Number of females recorded	Natural number
Male	Number of males recorded	Natural number
Worker	Number of workers recorded (for social species)	Natural number
Not.specified	Number of specimens recorded without sex identification	Natural number
Reference.doi	Digital object identifier	Text/link
Flowers.visited	Plant species where the specimen was collected	Text
Local.id	Unique identification in the local collection	Categorical
Authors.to.give.credit	All authors that have contributed on the original dataset	Text
Any.other.additional.data	Extra column for additional information	Text
Notes.and.queries	Notes or queries	Text
Unique.identifier	Unique identifier in the IberianBee database	Categorical

## FIGURE LEGENDS

**Figure 1** Workflow to build the ‘IberianBees’ database.

**Figure 2.** Heatmap of the number of bee records on the Iberian Peninsula. The records are in logarithmic scale with a gradient of colours. Areas with high number of records are coloured in yellow and areas with low number of records in purple.

**Figure 3.** Barplot of the number of records of the 20 most common species.

**Figure 4.** Number of records in the database by year (A) and month (B).

## PIES DE FIGURA

**Figure 1** Flujo de trabajo para construir la base de datos de ‘Iberianbees’.

**Figura 2.** Mapa de calor del número de registros de abejas en la península ibérica. Los registros se encuentran en escala logarítmica y se muestran con un gradiente de color. Zonas con alta densidad de registros aparecen en amarillo y zonas con una baja densidad en morado.

**Figura 3.** Gráfico de barras del número de registros de las 20 especies más comunes.

**Figura 4.** Número de registros en la base de datos por año (A) y por mes (B).

Figura 1

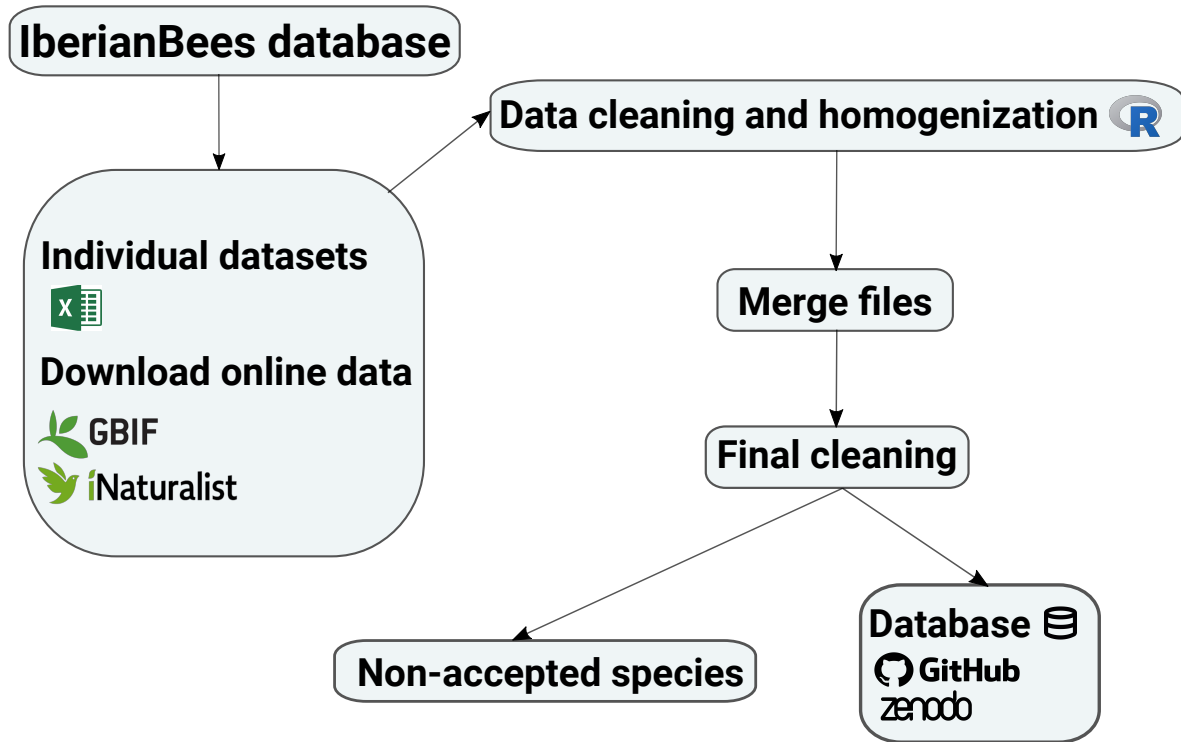


Figura 2

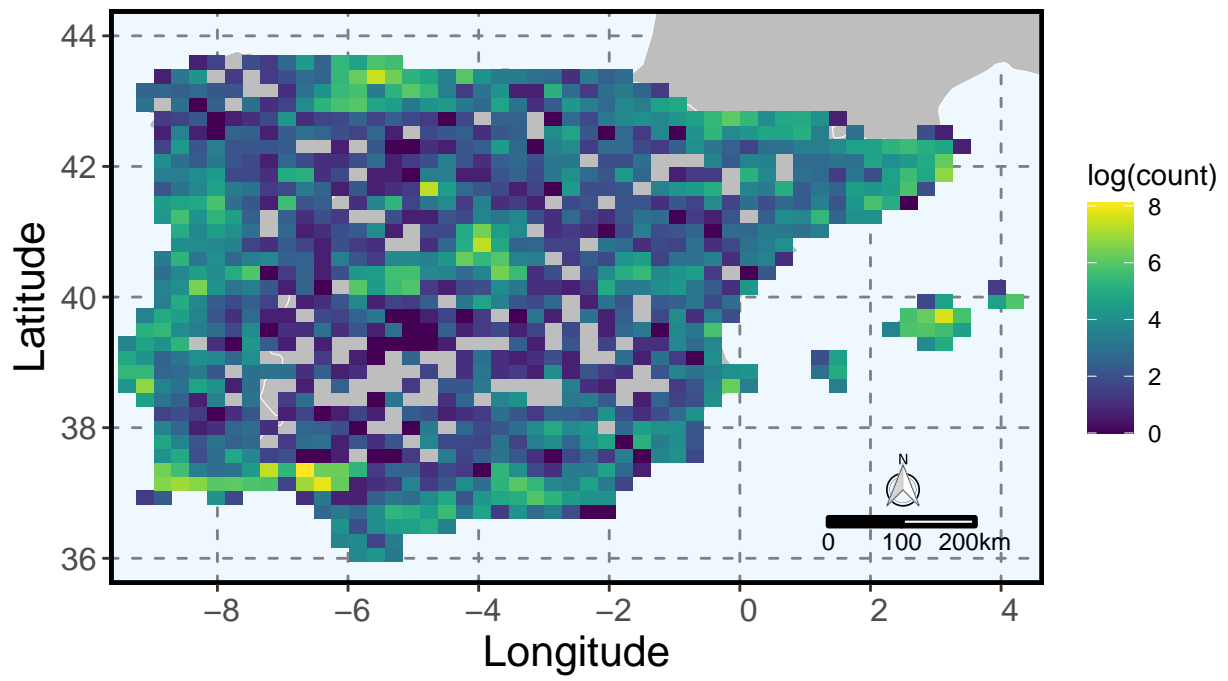


Figura 3

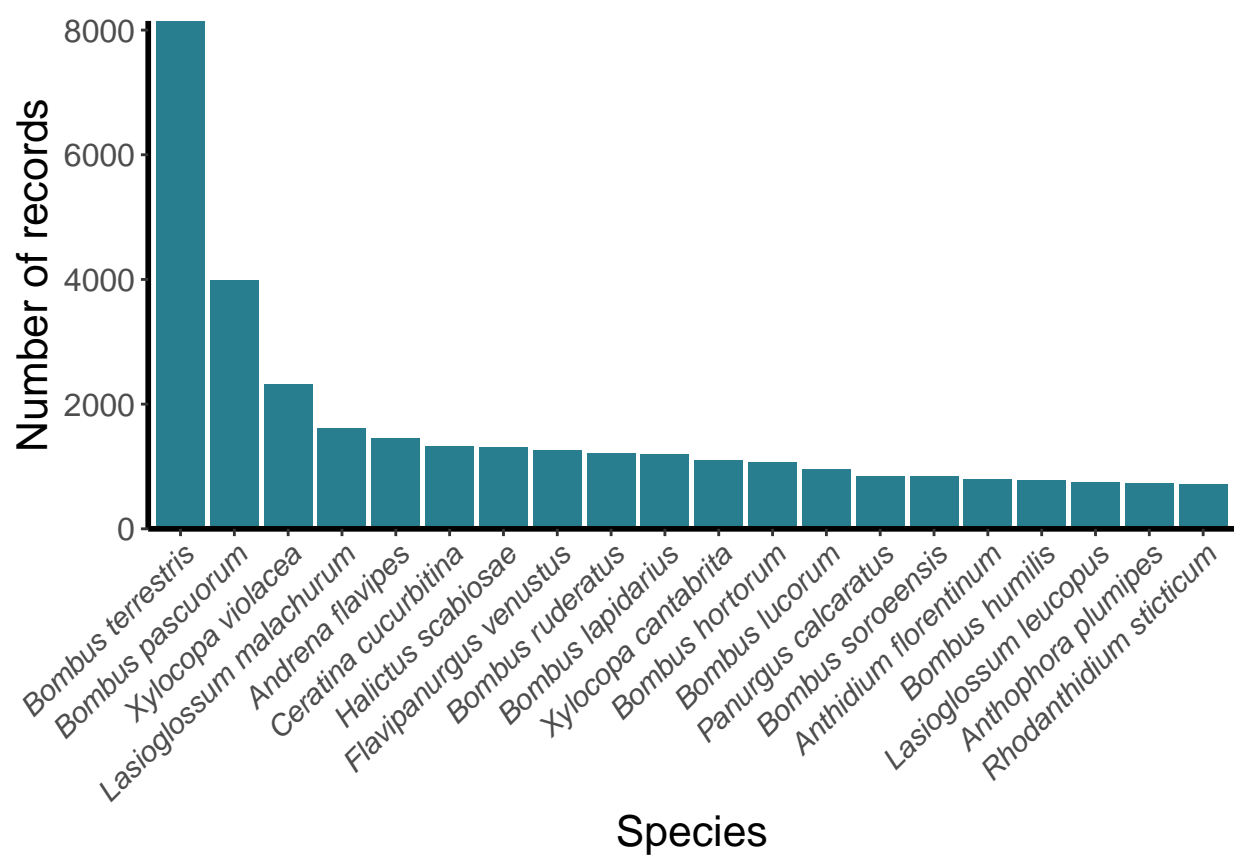


Figura 4

