Base de datos de abejas Ibéricas
Iberian Bees database

Ignasi Bartomeus[1], Jose B. Lanuza[1], Thomas Woods[2], Luisa Carvalheiro[3], and a lot of coauthors to be a

(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: https://doi.org/10.5281/zenodo.6354502

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: https://doi.org/10.5281/zenodo.6354502.

Palabras clave

Biodiversidad; Apoidea; Polinizadores; Fenología

Keywords

Biodiversity; Apoidea; Pollinators; Phenology

Background & Summary

Bees (Hymenoptera, Apoidea) are a diverse group of species with more than 20.000 species described worldwide (Ascher y Pickering, 2020). 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 (Sánchez, 2011). 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 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 of its populations (Goulson et al., 2015). 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 et al., 2013), other areas are well conserved (Herrera, 2019) and while some species are in the brink of extinction (Cameron et al., 2011) some species are even thriving (Russo, 2016). 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 exists. However, there are different sources of useful information. First, natural history museums hold historical collections of amateur naturalists and researchers who collected bees (Bartomeus et al., 2019). 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 where 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. Gbif) or collected by citizen scientists (e.g. iNaturalist). The challenge is how to digitalize, access, and curate all this different dara sources.

Here we compile an Iberian Bee Database of occurrence records by integrating digitalization efforts by leading national 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 its 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 depictted 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: https://doi.org/10.5281/zenodo.6354502.

Antecedentes y resumen ampliado

Las abejas (Hymenoptera, Apoidea) son un grupo muy diverso de especies con más de 20.000 especies descritas en el mundo (Ascher y Pickering, 2020). Debido a las condiciones climáticas de la Península Ibérica, esta se caracteriza por ser uno de los puntos calientes de diversidad a nivel mundial con más de 1000 abejas descritas hasta la fecha (Sánchez, 2011). A pesar de la gran diversidad de abejas Ibéricas, conocemos muy poco de su ecología y distribución. Paradójicamente las abejas son un un grupo clave que desarrolla una función vital para la moyoría de ecosistemaas terrestres, la polinización de miles de plantas (Ollerton et al., 2011) incluyendo a la mayoría de cultivos (Klein et al., 2007). Además el cambio debido al impacto del ser humano esta dando lugar al declive de algunas de sus poblaciones (Goulson et al., 2015). No obstante, la respuesta de las abejas al cambio global es heterogénea tanto en el espacio como taxonomicámente. 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 mientras que algunas especies se encuntra al borde de la extinción (Cameron et al., 2011), otras se encuentran incluso prosperando (Russo, 2016). Una mejor descripción de la distribución de las abejas sobre su distribución, preferencia de hábitat, fenología, o tendencias históricas en la Península Ibérica requerirían información de dónde y cuándo las especies aparecen.

Con la ausencia de esfuerzos de muestreo estandarizados se podría pensar que información detallada de la ocurrencia de especies en el espacio y en el tiempo no existe. Sin embargo, existen diversas fuentes con información muy útil. En primer lugar, los museos de historia natural albergan colecciones históricas de naturalistas amateur e investigadores que coleccionaron abejas (Bartomeus et al., 2019). Puesto que las colecciones de museos suelen estar muy bien clasificadas y etiquetadas, es posible recopilar la información sobre los especímenes de la colección y cuándo y dónde estos fueron capturados. En segundo lugar, la Península Ibérica tiene una rica tradición de investigadores sobre polinizadores y polinización (Archer et al., 2014) que han recopilado gran cantidad 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). No obstante, sigue siendo un desafio digitalizar, acceder y limpiar toda la información existente de las diferentes fuentes de datos.

En este estudio hemos creado una base datos de abejas Ibéricas (IberianBees) a través de unificar: (i) los esfuerzos de digitalización de museos, (ii) conjuntos de datos de investigadores, e (iii) información de repositorios en línea de libre acceso. Tras la limpieza de los datos y homogeneización de los mismos para crear la base de datos, nosotros liberamos los primeros 103464 registros que abarcan la Península Ibérica e Islas Baleares. La **Figure 1** muestra la distribución espacial de los registros, tal y como se esperaba existe un sesgo hacia ciertas áreas cercanas a los grandes núcleos urbanos. Esta base de datos contiene información sobre 925 diferentes especies y las 20 especies más comunes se encuentran representadas con su respectivo número de registros en la **Figura 2**. Los datos han sido recopilados entre `rmin(data$Year, na.rm = T)` y 2022, con un ascendente número de registros en el tiempo que refleja el creciente interés de documentar a este grupo taxonómico (**Figura 3A**). Además, como es de esperar la mayoría de registros tienen lugar en los meses de primavera con mayo siendo el mes que consta de un mayor número de registros (**Figura 3B**). El 89% de estos registros constan de información georeferenciada y el resto de registros al menos contienen información de la localidad de captura/observación. También un 55% de los diferentes registros poseen información sobre la fecha de captura y para los que no, un 1% tienen un intervalo temporal de cuándo se pudieron capturar/observar los diferentes especímenes. Además para un 23% de los registros hay información de la(s) planta(s) donde se capturo o se produjo la observación. Esta es la versión (v.1.0) de la base de datos de abejas Ibéricas con nombre "IberianBees" que continuará creciendo pero creemos que es un buen punto de partida para empezar a mejorar nuestro conocimiento y conservación de este grupo taxonómico tan importante para nuestros ecosistemas y la producción agrícola. Se puede acceder desde: https://doi.org/10.5281/zenodo.6354502.

Material y métodos

Recopilación de datos originales

Museos: Hemos digitalizado el 10% de los espcimenes depositados en el Museo Nacional de Ciencias Natural de Madrid, sobretodo 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, asi como anunciado en las principales listas de distribución tematicas para recoger datos sobre ocurrencia de abejas en la Peninsula Iberica. Se recogieron 157 estudios con datos ya digitalizados. Estos han sido limpiados y homogeneizados (ver más abajo). En el futuro se seguiran añadiendo nuevas contribuciones, asi como se rescatarán datos ya publicados en formatos de dificil extracción como PDF.

Repositorios de internet: Usando paquetes programaticos de R ((**ropensciINat?**): https://cran.r-project.org/web/packages/rinat/rinat.pdf, (**ropensciGbif?**): https://cran.r-project.org/web/packages/rgbif/citation.html) descargamos todos los datos de Gbif (https://www.gbif.org) y INaturalist (https://www.inaturalist.org) de ocurrencias de abejas en la peninsula Iberica a fecha 14-03-2022. Estos datos pueden ser actualizados en nuevas versiones automaticamente.

Tratamiento de datos

Todo el tratamiento de datos se ha hecho de forma reproducible en R (R Core Team, 2022). Primero hemos seleccionado los campos comunes que reflejan que especie y de que sexo, cuando fue colectada, y donde, asi como datos de quien fue el colector, quien determinó la identidad de la especie, y referencias de donde ha sido publicada. La Tabla 1 ofrece un resumen de los metadatos creados con (Boettiger et al., 2022), y que también están disponibles en formato json (indexado por Google datasets) y (Jones et al., 2019). Debido a la naturaleza variada de este tipo de datos el proceso de limpieza de datos fue un proceso complejo y tedioso. De los 108140 datos iniciales, 4676 fueron descartados. Todo el proceso así como los datos descartados pueden consultarse aqui: https://github.com/ibartomeus/IberianBees

En primer lugar se reviso toda la taxonomia de las especies cotejándolos con la lista de especies actualizada para la Península Ibérica (Sánchez, 2011). 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 tipográficos. En segundo lugar, se ha recogido información del país, provincia y localidad del registro. Además se realizó un proceso de estandarización de las geolocalizaciones a latitud y longitud en grados decimales (WGS84) siempre y cuando esta información se encontraba disponible. Aunque también se incluyó información sobre la precisión de las coordenadas, este metadato estaba raramente presente para muchas de las coordenadas. En tercer lugar, se ha incluido también información de la fecha de captura/observación con el día, mes y el año o en su defecto con intervalo temporal orientativo si esto era posible. Los datos con fechas y localizaciones erróneas o ilógicas (e.g. geolocalizaciones en el mar o especimeénes recogidos en el mes 18) han sido corregidos o eliminados. En cuarto lugar, se ha recogido los nombres de los colectores y personas que identificaron las especies, así como información del 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 unica de la base de datos local (si la hay), los autores que han proporcionado los datos y cualquier otra nota de interés si la había.

En la base de datos resultante se ha asignado a cada entrada un identificador unico, y se ha guardado tanto el nombre de la espcecia original tal y como fue inicialmente identificada, como el nombre aceptado final después de haber sido revisado por taxónomos expertos si era necesario.

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, 2022) y Rmarkdown (Xie, 2014, 2015, 2021; Xie et al., 2018, 2020; Allaire et al., 2022) para todos nuestros análisis.

Registro y disponibilidad de datos

Todo el proceso de trabajo esta en abierto en Github y los todos los archivos del código creado tienen licencia MIT (https://opensource.org/licenses/MIT). El uso de datos tine licencia CC-By 4.0 (https://creativecommons.org/licenses/by/4.0/). La versión depositada junto al envio de este paper corresponde a la version 1.0. de IberianBees, y tanto los datos como los scripts tienen una version permanente depositada en Zenodo y citable usando este identificador único. Se usará un sistema de versionado incremental, donde pequeñas correcciones de errores, o incrementos no substanciales de datos recibiran actualizaciones en el primer digito (v1.1, v.1.2, etc...) y actualizaciones mayores en el segundo numero (v2.0, v3.0, etc...). Todas las actualizaciones mayores serán depositadas en Zenodo bajo el DOI general: https://doi.org/10.5281/zenodo.6354502

Contribución de los autores

IB and JBL 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 JBL 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. 2022. *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.

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.

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: 1255957.

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.

Ollerton, J., Winfree, R., Tarrant, S. 2011. How many flowering plants are pollinated by animals? *Oikos* 120: 321-326.

R Core Team. 2022. *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.

Sánchez, F.J.O. 2011. Lista actualizada de las especies de abejas de España (Hymenoptera: Apoidea: Apiformes). *Boletín de la Sociedad Entomológica Aragonesa*265-281.

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. 2021. *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.

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
Longitude	Longitude	Decimal degrees
Coordinate.precision	Precision at which the coordinates are provided	Categorical
Year	Year of collection	Year
Month	Month of collection	Month

variableName	description	unitText
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 publiched, 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	List of authors providing the original dataset	Text
Any.other.additional.data	Free text to note any other additional information	Text
Notes.and.queries	Free text to note any other notes or queries	Text
uid	Unique identification in IBD.	id
original_name	Species name as provided by the orignal dataset	Categorical
accepted_name	Final species name after correcting for misspellings and synonims	Categorical

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 moths.

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.

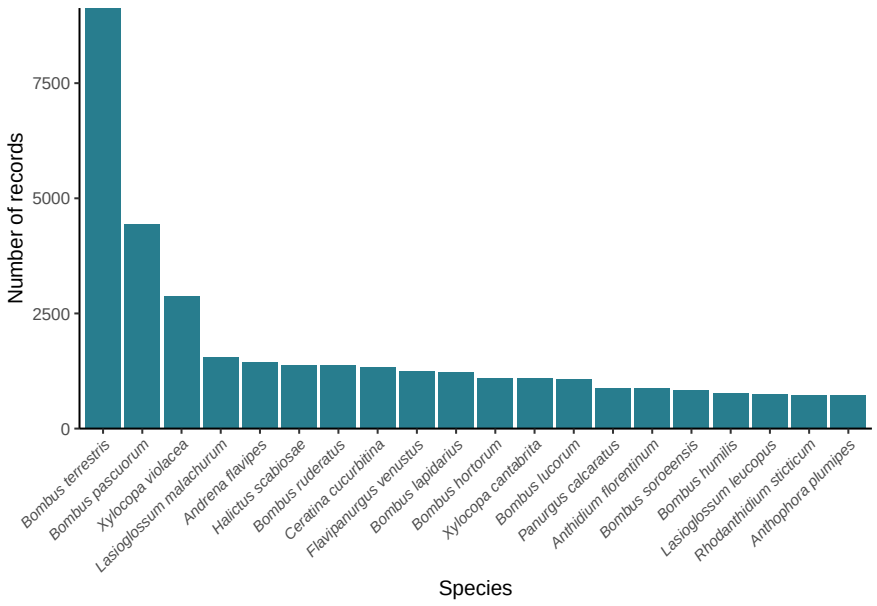

Figure 1: Barplot indicating the 20 more common recorded species in the final data set.

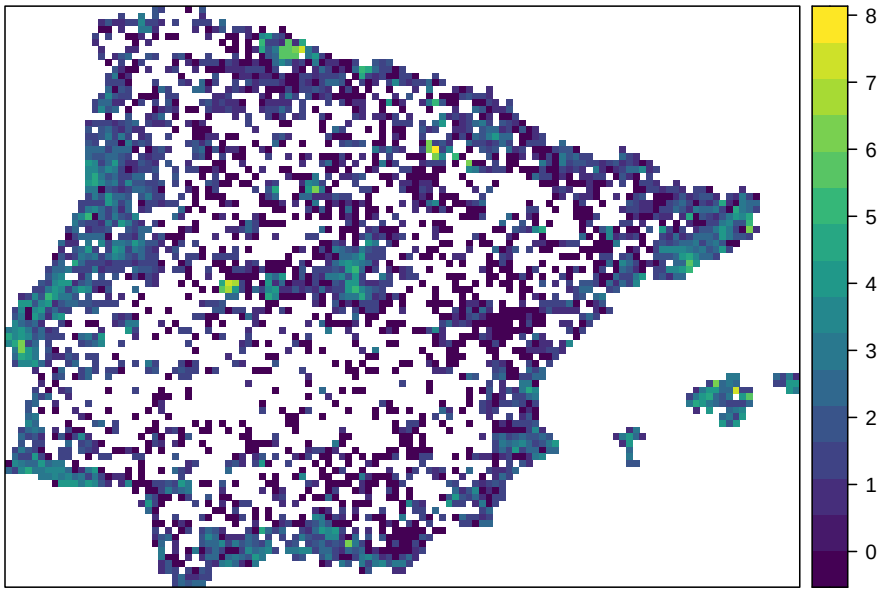

Figure 2: Map of the Iberian peninsula indicating where the bee occurrences were sampled.

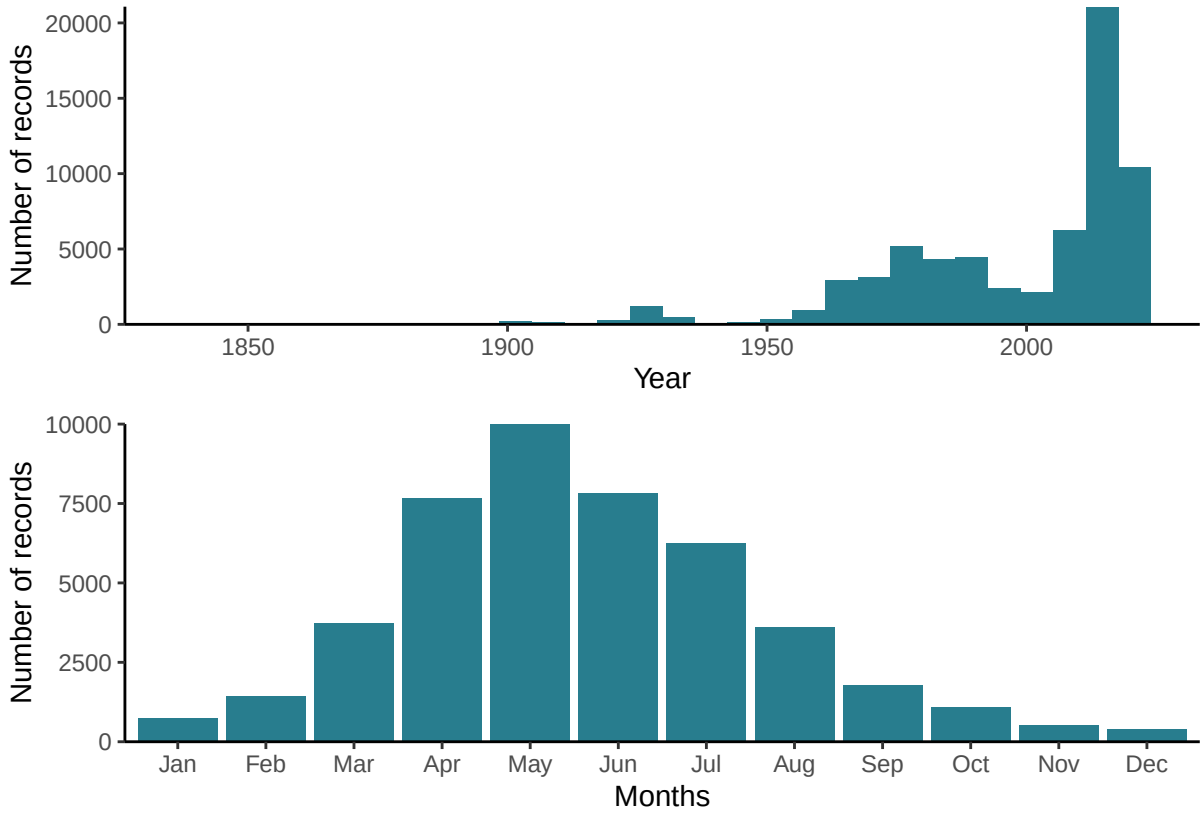

Figure 3: Cobertura temporal de (A) los años con más ocurrencias, y (B) los meses con más ocurrencias.