

Are patterns of sampling effort and completeness of inventories congruent? A test using databases for five insect taxa in the Iberian Peninsula.

David Sánchez-Fernández¹, José Luis Yela², Raúl Acosta³, Núria Bonada³, Enrique García-Barros⁴, Cástor Guisande⁵, Juergen Heine⁶, Andrés Millán¹, Miguel L. Munguira⁴, Helena Romo⁴, Carmen Zamora-Muñoz⁷, Jorge M. Lobo^{8*}

¹Departamento de Ecología e Hidrología, Universidad de Murcia, Murcia, Spain

²DITEG Research Group, Facultad de Ciencias Ambientales y Bioquímica, Universidad de Castilla-La Mancha, Toledo, Spain.

³Freshwater Ecology, Hydrology and Management group (FEHM), Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain.

⁴Departamento de Biología y Centro de Investigación en Biodiversidad y Cambio Global (CIBC-UAM), Universidad Autónoma de Madrid, Spain.

⁵Facultad de Ciencias, Universidad de Vigo, Vigo, Spain.

⁶Department of Computer Science, Universidad de Vigo, Vigo, Spain.

⁷ Departamento de Zoología, Universidad de Granada, Granada, 18017, Spain.

⁸ Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain.

*corresponding author: jorge.lobo@mncn.csic.es

Full postal address: Jorge M. Lobo. Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, CSIC, Madrid, Spain. C/ José Gutiérrez Abascal 2, 28006 Madrid (Spain).

Telephone number: +34 914 11 13 28.

Keywords

Iberian Peninsula, aquatic beetles, dung beetles, butterflies, moths, caddisflies, biodiversity inventories, insect decline.

Running title: low congruence in insects inventorying

35 **Abstract**

- 36 1. Evaluating data quality and inventories completeness must be a preliminary step in
37 any biodiversity research, particularly in the case of insects and high biodiversity
38 areas. Yet, this step is often neglected or, at best, assessed just for only one insect
39 group, and the degree of congruence for different insect groups of sampling effort
40 remains unexplored.
- 41 2. We assess the congruence in the spatial distribution of the sampling effort for five
42 insect groups (butterflies, caddisflies, dung beetles, moths, and aquatic beetles) in the
43 Iberian Peninsula. We identify well-surveyed areas for each taxonomic group and
44 examine the degree to which the patterns of sampling effort can be explained by a set
45 of variables related to environmental conditions and accessibility.
- 46 3. Irrespective of the general lack of reliable inventories, we found a general but low
47 congruence in the completeness patterns of the different taxa. This suggests that there
48 is not a common geographical pattern in survey effort and that idiosyncratic and
49 contingent factors (mainly the proximity to the workplaces of entomologists) are
50 differentially affecting each group.
- 51 4. After many decades of taxonomic and faunistic work, distributional databases of
52 Iberian insects are still in a very preliminary stage, thus limiting our capacity to obtain
53 reliable answers to basic and applied questions.
- 54 5. We recommend carrying out long-term, standardized and well-designed
55 entomological surveys able to generate a reliable image of the distribution of different
56 insect groups. This will allow us to estimate accurately insect trends and better
57 understand the full extent of global biodiversity loss.

58

59

Introduction

Understanding how biodiversity is distributed in space and time is one of the main challenges for ecologists, biogeographers and conservation biologists (Cox & Moore, 2004). Reliable species distributions data provide key information to understand, not only the causes behind current biodiversity patterns (Gaston, 2000) and future biodiversity loss, but also to guide efficient conservation planning strategies (Meyer *et al.*, 2015; Kujala *et al.*, 2018).

Insects make up the bulk of terrestrial biodiversity (Stork, 2018). They are a central component of the living world, and their protection is crucial to maintaining ecosystems and ensuring food security (Hallmann *et al.*, 2017; Bélanger & Pilling, 2019). It is therefore difficult to address a global study of biodiversity, or its conservation, without taking into account this group. However, their hyperdiverse character and the general scarceness and biases of the available information (Hortal *et al.*, 2015) make this endeavour difficult. Even though much of the scientific attention has focused on the evidence for and effects of declining insect biodiversity (Goulson, 2019; Habel *et al.*, 2019; Cardoso *et al.*, 2020; Wagner *et al.*, 2021), other studies have cautioned against over-extrapolating this current evidence (Saunders *et al.*, 2020) because of the heterogeneous responses of insects and lack of comparable inventories (Wagner, 2020). Indeed, the phenomenon of insect decline is generally poorly understood for many insect groups, and there is an urgent need for more reliable data to estimate the true extent of global and regional insect decline.

Despite the widely recognized importance of evaluating data quality and inventories completeness as a preliminary step in any biodiversity study, this process is often neglected. Only countries with a long-standing tradition of natural history and sufficient resources can produce good databases and distribution maps based on large volumes of data (Lawton *et al.*, 1994; Griffiths *et al.*, 1999; Sánchez-Fernández *et al.*, 2021). Unfortunately, this is usually not the case for countries located in the Mediterranean Basin, which despite being included in one of Earth's biodiversity hotspots (Myers *et al.*, 2000), are characterized by incomplete, unreliable or non-existent insect inventories (Ramos *et al.*, 2001, but see for example Ruffo & Stoch, 2006). This lack of information can be considered the main factor limiting both the study of biodiversity patterns in these countries and the identification of priority areas for the conservation of insects.

The examination of faunistic databases carried out to date in the Mediterranean Basin have reported a dearth of complete and extensive inventories for several insect groups (e.g., Romo *et al.*, 2006; Sánchez-Fernández *et al.*, 2008; Bruno *et al.*, 2012; Fattorini, 2013; Lobo *et al.*, 2018). Within this region, the Iberian Peninsula is an area with an outstanding

biodiversity and a high level of endemism (Rosso *et al.*, 2018) due to the climatic conditions during Late Quaternary and its role as glacial refuge (see Gómez & Lunt, 2004; Schmitt, 2007; Abellán & Svenning, 2014). The few studies conducted so far in the entire Iberian Peninsula (Romo *et al.*, 2006; Sánchez-Fernández *et al.*, 2008; Lobo, 2008; Lobo *et al.*, 2018) have identified significant geographic data gaps for individual insect groups, but also a well-defined spatial and environmental pattern in the temporal accumulation of Iberian dung beetle records (Lobo *et al.*, 2007). However, an examination of both the biases and patterns in the data collection carried out among different insect groups is lacking so far. One may find congruence in the spatial and environmental characteristics of the survey effort made by the naturalists interested in different insect groups because of their common preference for some attractive places such as protected areas, mountains, accessible places or localities near research centres and universities (Romo *et al.*, 2006; Sánchez-Fernández *et al.*, 2008; Mair & Ruete, 2016). If this is the case, we might be more likely to find a general picture of the distribution of insect survey effort in the Iberian Peninsula, and as a result design shared sampling campaigns to overcome data gaps. On the contrary, a lack of congruence among the different geographical patterns of survey effort may arise from specific and contingent factors related with collectors' preferences (Dennis & Thomas, 2000) or repeated surveys in localities characterized by the presence of rare or interesting species (Sastre & Lobo, 2009). In this latter case, it would be difficult to obtain general insect patterns of data accumulation in the short to medium term, and survey effort should be specific and more expensive.

Despite the potential importance of this topic, to our knowledge, there are no studies assessing the congruence in the patterns of survey effort (completeness) using databases of different insect groups. Our general aim is to assess the congruence in the distribution of the sampling effort carried out for five insect groups in the Iberian Peninsula (true aquatic beetles, dung beetles, butterflies, moths and caddisflies). Our aims were: i) to identify the most probable well-surveyed areas for each taxonomic group, and ii) to examine the extent to which sampling bias can be explained by a suite of environmental and accessibility related variables.

Methods

Data source

This work is based on five up-to-date and exhaustive databases (Table 1) of five taxonomic groups of insects: caddisflies (Trichoptera), noctuid moths (Lepidoptera: Noctuoidea), dung beetles (Coleoptera: Geotrupidae, Scarabaeinae and Aphodiinae), aquatic beetles (Coleoptera:

Dryopidae, Dytiscidae, Elmidae, Gyrinidae, Haliplidae, Helophoridae, Hydraenidae, Hydrochidae, Hydrophilidae, Hygrobiidae, Noteridae and Sphaeriidae), and butterflies (Lepidoptera: Papilionoidea) from the Iberian Peninsula and Balearic Islands. The five databases mentioned above compile available georeferenced species records from the literature, as well as records from museums and private collections, PhD theses, field records, and various type of grey literature. These databases started some time ago and their first versions were published in some cases (Lobo & Martín-Piera, 1991; García-Barros *et al.*, 2004; Millán *et al.*, 2014) and included in GBIF (www.gbif.org/). However, current versions include a large number of new and revised records. Doubtful records (taxonomically uncertain or imprecisely georeferenced) were deleted by experts on each taxonomic group, resulting in 738,496 species records (with caddisflies having the least number (13,961), and butterflies having the most (423,820)). The five databases cover a total of 2,257 species, ranging from the least diverse group, butterflies, with 230 species to the most diverse, moths, with 848 species (see Table1). All these databases will be freely accessible within the GeoBrink platform (<http://geobrink.uclm.es>)

Measuring data quality

Data analyses were conducted using as spatial units the 353 cells with a resolution of 25 minutes, which means an approximate cell size of 36 x 46 km at the average latitude of the Iberian Peninsula..

A computer application called KnowBR (Lobo *et al.*, 2018) was used to examine the reliability of biodiversity inventories in each one of the 353 cells for each taxonomic group. To do this, KnowBR uses “species accumulation curves”, a general approach in which the number of collected species is related with the increase in survey effort (Soberón & Llorente, 1993; Colwell & Coddington, 1994). This application automatically estimates the final slope of the accumulation curve for each cell, the completeness of each cell inventory, and the ratio between the number of database records and the number of species in each cell.

Accumulation curves were calculated after adjusting the data to a linear rational function (see Flather, 1996 for details) which is simply the ratio of two polynomial functions. Rational functions have excellent asymptotic properties being able to represent an extremely wide range of shapes (Bolker, 2008). The asymptotic values were estimated for all these curves to extrapolate the probable number of species in each cell when the number of records tend towards infinity. The percentage of species that have been recorded over these predicted asymptotic values is the completeness value. Completeness values and the final slope of these

relationships were then used as surrogates for the quality of the survey effort carried out (Lobo *et al.*, 2018). When these two parameters are derived from different databases, the slope at the end of the accumulation curve and the completeness values are highly correlated (Spearman rank correlation $r_s = -0.981$, $P < 0.0001$; Fig. S1 in Supplementary Material). However, similar low slope values can be attributed to the cells with more than 80% of completeness values due to the curvilinear relationships between these two parameters (see Fig. S1). Consequently, we used completeness figures derived from accumulation curves as the best indicator to measure the reliability of cell inventories.

We measured the quality of cell inventories using six ordinal categories (Table 1): *ignorance cells*, *very poorly sampled*, *poorly sampled*, *moderately sampled*, *relatively well-sampled* and *well-sampled*. Ignorance cells are those without any record. Very poorly sampled cells are those with some records but too few to calculate an asymptotic or quasi-asymptotic curve, as the relationship between the number of collected species and the number of database records is almost linear. The mean ratio between the number of database records and the number of species in these cells is 1.4 ± 2.3 (SD). Poorly sampled cells were identified as those with completeness values greater than 1% but less than 50%. Moderately sampled cells have completeness values from 51% to 75%, while relatively well-sampled cells have completeness values from 76% to 90%; and well-sampled cells are those having completeness values higher than 90%.

Predictor variables

Twelve environmental variables were considered to explain the variation in completeness across the geography for each insect group. We used five main climatic variables, annual mean temperature, maximum temperature of the warmest month, minimum temperature of the coldest month, annual precipitation, and precipitation of the driest month, derived from WorldClim v. 2.1 (www.worldclim.org; Fick & Hijmans, 2017). We also used as predictors the amount of terrestrial area not covered by marine water and the mean elevation of each 25 minute cell. Elevation was calculated from a digital elevation model (DEM) downloaded from the USGS EROS Data Center (<https://www.usgs.gov/centers/eros>). The different land uses of level 2, from CORINE Land Cover 2018 (<https://land.copernicus.eu>), were reclassified into four main categories: anthropic areas (including artificial surfaces and agricultural areas), forests, shrubs, and grasslands. The area of each one of these four land cover categories was calculated for each cell square. Finally, mean soil pH data for each cell was obtained from the European Soil Data Centre (<http://esdac.jrc.ec.europa.eu/>; see Reuter *et al.*, 2008).

In addition to these environmental variables, a predictor related to accessibility was calculated. Based on their knowledge about the historical development of the studies in their group and the distribution of the number of records attributed to the main taxonomists, collectors and/or research centres, researchers of each insect group subjectively selected three main localities that acted as study centres. Once the localities were selected, the distance from the central point of each cell to the nearest locality (out of the three) was used as a predictor to account for a bias in the recording effort carried out in each insect group (see Fig S2 in Supplementary Material). The number of localities selected for each distance map was a trade-off between simplicity and complexity in the distance maps generated according to the selected number of study centres.

Data analysis

To analyse spatial congruence, completeness values for all the taxonomic groups in the 25 minute cells were submitted to a Principal Component Analysis (PCA) in order to discriminate the main orthogonal variables with eigenvalues higher than one, thus reflecting the comparative spatial variability among groups in the survey effort. The same completeness data were also analysed with a Cluster Analysis to better visualise the relationships between insect groups. Ward's method was used as linkage rule and 1-Pearson r as the distance measure to maximise obtaining clusters with similar numbers of objects (Legendre & Legendre, 2012).

We used Generalized Linear/Nonlinear Models (GLMs; see McCullagh & Nelder, 1989) to quantify the individual explanatory capacity of each predictor on the presence-absence (1/0) of "reliable" inventories in 25 minute cells. Unreliable cells (0) were those without records or with very few records that it was impossible to estimate the accumulation curve (ignorance and very poorly sampled cells). Reliable cell inventories (1) were delimited as those belonging to the remaining four cell categories in which it was possible to estimate the predicted richness and completeness values from accumulation curves (poorly sampled, moderately sampled, relatively well-sampled and well-sampled). A binomial error distribution linked to the set of predictors via a logit link function was used in this case. Subsequently, continuous completeness values in cells with reliable inventories were used to run another GLM to examine the predictors able to explain the completeness variability in the surveyed cells. A normal distribution was assumed and a logarithmic link function was used in this case. The two statistical approaches allowed us to examine whether the explanatory variables discriminating between cells with and without informative species inventories were

similar to those explaining degrees of completeness. The goodness-of-fit of the model was measured by the statistic deviance calculating the percentage of change in deviance from a null model in which no predictors were included. The Wald statistic was used for testing the statistical significance of the regression coefficients based on maximum likelihood estimates (see McCullagh & Nelder, 1989). All the predictors were standardized (to a mean of zero and standard deviation of one) to eliminate the effect of the measurement scale differences. All these statistics were performed using StatSoft's STATISTICA v10.0 (StatSoft Inc., Tulsa, Oklahoma, USA).

Results

The number of cells considered to be well sampled (category 6) represented less than 4% of the total study area (from 1 to 11 cells depending on the group), except in the case of butterflies (96 cells; 27.20%), the group with the highest number of records per species (1,843; Table 1; Fig. 1C). Caddisflies was the group least surveyed, in which almost 69% of the cells were assigned as very poorly sampled or ignorance areas (categories 1 and 2; Table 1, Fig. 1A). Moths, dung beetles and aquatic beetles shared similar results, with just over half of their cells categorized as moderately sampled and relatively well sampled (categories 4 and 5) (Table 1; Fig. 1).

Two PCA axes with eigenvalues (>1) were selected, accounting for 53.0% of the completeness variation (29.8% and 23.2%, respectively) for the five insect groups. Both the factor loadings and the Cluster Analysis allowed us to separate these five groups into two main clusters (Fig. 2): one joining the two Lepidoptera groups and another including the remaining insect groups in which dung beetles were placed in an intermediate position. The distribution of the six ordinal categories in the cells for each insect group showed that the quality of the inventories was partially congruent between butterflies and moths (Spearman rank correlation $r_s = 0.274$, $p < 0.0001$). Well-surveyed cells for these two groups were located in north-eastern Iberian Peninsula, while the worst surveyed cells were mainly distributed across the central and northwest areas of the Iberian Peninsula (Figs. 1C and E). The distribution of the six ordinal categories in caddisflies and aquatic beetles were also partly congruent ($r_s = 0.232$, $p < 0.0001$). Wide areas of central and south-western Iberia remained poorly surveyed for both groups, while some cells of the southern Mediterranean area of the Peninsula were well surveyed. The case of dung beetles was relatively singular

(Fig. 1D), with cells with different quality values in their inventories distributed all over the study area.

The cells with poor surveys (categories 1 and 2) seem to be preferably located in the central part of the Peninsula, irrespective of the group (see Fig. 1); 27 cells belong to these categories for at least four insect groups (Fig. 3A). On the other hand, the best surveyed areas, (categories 5 and 6) seem to be located in the main mountainous systems of the Iberian Peninsula (Pyrenees, Cantabrian mountains, the Iberian Central System, and along the Penibaetic System (SE Spain)) (Fig. 1F). However, only two cells are considered as belonging to these categories for at least four groups (Fig. 3B), and large areas of the Ebro valley (north-west) and the southern plateau in Central Spain do not harbour well-surveyed cells for any group of insects (Fig. 3B).

The explanatory capacity of the selected predictors on the presence-absence of reliable inventories oscillates from 8.2% in the case of dung beetles to 24.4% for caddisflies (Table 2). Elevation and soil pH appear as the most relevant variables, always with positive signs in their coefficients. Thus, cells with a lower elevation and with more acidic soils have a higher likelihood of remaining either not surveyed or poorly surveyed (Table 2). Maximum temperature of the warmest month and the distance to the main research centres are also relevant variables in some groups with negative signs, i.e., not surveyed or poorly surveyed cells tend to be located in areas with warmer summers and away from research centres (Table 2). Including the distance to the main research centres in the models with only the environmental variables gives a 15.6% average increase in the percentage of explained variability (max = 60.4%; min = 0.2%).

The explained variation of the selected predictors in the case of continuous completeness values in the cells with reliable inventories oscillated from 6.5% in aquatic beetles to around 22% in dung beetles and caddisflies (Table 2). Environmental variables were relevant to explain the geographical pattern of completeness for only two insect groups. In general, completeness is higher when the cell elevation and the cell mean temperature increase in butterflies and dung beetles (Table 2). However, the most outstanding result is related with the role played by the variable representing the distance to the main research centres. In all the considered groups, this simple variable appears as statistically significant, always with negative sign. Thus, places located far away from these research centres have biological inventories characterized by low completeness values. Models including this distance predictor compared with those only including environmental variables increase substantially the explained

deviance in moths (88%) and dung beetles (50%), but also in aquatic beetles (33%), caddisflies (29%) and butterflies (20%) (mean = 44.1%; Table 2).

Discussion

Our results show that knowledge on the geographical distribution of insect biodiversity in the Iberian Peninsula is poor and geographically biased. This indicates that the Wallacean shortfall clearly emerges in the Mediterranean Basin (Brown & Lomolino, 1998; Whittaker *et al.*, 2005), most likely one of the best studied biodiversity hotspots in the world (Blondel & Aronson, 1999). Despite the wide size of our spatial units (25 minutes), large areas remain poorly surveyed. The lowlands of the central Iberian Peninsula appear to be particularly under-surveyed, perhaps as a consequence of the secular degree of anthropogenic alteration of these landscapes and/or their unattractive character for entomologists. Interestingly, when the new terrestrial animal species described in the Ibero-Balear region from 1994 to 2012 are mapped, no environmental and land use predictors are able to explain the characteristics of the localities in which the species have been found, and a large number of new species may be discovered in a locality when it is subjected to intense survey effort (Payo-Payo & Lobo, 2016). Therefore, we suspect that the lack of survey effort in the lowlands of the central Iberian Peninsula is largely due to the preferences of insect collectors.

The biased character of the geographical knowledge about Iberian insects is especially worrying if we consider that the five insect groups studied are among the groups for which we have a better knowledge from a faunistic point of view. We must stress the difficulty to get complete inventories for insects. For example, Sánchez-Fernández *et al.*, (2021) recently analyzed the database of Great Britain butterfly occurrences, likely the most complete database for any insect group in the world (with more than 10 million records for 58 butterfly species), and found that there are still some areas in Great Britain without complete inventories. The best way to circumvent this Wallacean shortfall is to invest in biodiversity sampling programs (Balmford & Gaston, 1999) or, a less preferable alternative, attempt to develop species distribution models (Guisan *et al.*, 2017) providing reliable geographical predictions. However, to make these modeling procedures realistic, it is essential to have a minimum of cells with reliable inventories representing the environmental and spatial spectrum of the study area (Hortal & Lobo, 2005).

We can establish three sets of taxonomic groups based on the completeness of their inventories in Iberia: a very poorly studied group, in which more than 68% of the considered cells do not have data or have very few samples (caddisflies), an intermediate group made up

of moths, dung beetles and aquatic beetles, in which the percentage of poorly sampled cells fall around 30%, and a much better sampled group (butterflies) with many records and fewer species richness.

Irrespective of the lack of reliable inventories for insects and the disparities in the completeness values of the studied groups, we found a general low congruence among the different completeness patterns. It is worth noting that the congruence is higher for those groups that are taxonomically related (i.e., butterflies and moths, or aquatic and dung beetles) or share the same type of habitats (i.e., freshwater habitats by aquatic beetles and caddisflies). Anyway, this general lack of congruence suggests that there is not a common geographical pattern in the survey effort to known insect distributions, and that idiosyncratic and contingent factors are differentially affecting each group (see a similar result in Mair & Ruete, 2016 for citizen biodiversity data). Although the percentages of explained variation are quite different between the studied groups, our results indicate that the selection of the areas to be sampled at least once is mainly determined by environmental factors as in other cases (Lobo *et al.*, 2007; Tiago *et al.*, 2017; Speed *et al.*, 2018). However, the differences in survey effort within these sampled areas seem to be more associated with sociological factors, such as the distance to research centres (see Romo *et al.*, 2006; Dennis & Thomas, 2000). Thus, entomologists have surveyed more intensely mountain areas (with alkaline soils), but the differences in survey effort within these areas could be conditioned by the proximity to workplaces (i.e., trying to decrease the fieldwork investment) (Oliveira *et al.*, 2016). Nevertheless, it must be noted that the predictors used cannot explain almost three quarters of the variation in completeness values. The low spatial resolution of our analyses could be behind this result, but also the lack of a clear environmental determinism and the randomness of the survey decisions. Other unique factors not considered here could also be relevant for explaining the geographical pattern of survey effort. Of these, the following should be underlined: the location researchers' second residence, their taxonomic curiosity (Sastre & Lobo, 2009), the existence of specific research programs on protected areas, or the presence of locations in which species new to science have been described (Tulloch *et al.*, 2013; Payo-Payo & Lobo, 2016).

The provided results have deep implications in conservation, such as to estimate accurately insect trends and to better understand the full extent of global biodiversity loss. Thus, it is essential to perform an assessments scheme specific to groups and the completeness of their inventories and the factors that may explain them. Hence, our inability to provide reliable snapshot estimations of the distribution of a large portion of biodiversity could be preventing us from taking advantage of the available occurrence data for basic and applied

purposes. The consequences of declines in insect biodiversity (Cardoso et al. 2020) can be unpredictable, and call for immediate policy responses (Forister et al. 2019; Harvey et al. 2020). However, after many decades of taxonomic and faunistic work, our biodiversity databases are still preliminary in many different aspects, and far from being able to provide useful information to solve basic and applied guidance by themselves, i.e., we are still unable to discriminate areas with low richness from those that simply are badly sampled. Novel and integrative applications should thus be restricted to certain taxonomic groups and regions with higher numbers of quality records (Ball-Damerow *et al.*, 2019). New extensive sampling fields (from long-term standardized and well-designed entomological surveys; Montgomery *et al.*, 2021), continued data digitization, publication, enhancement, and quality control efforts will continue to be necessary to make biodiversity science more efficient and relevant for conservation purposes. Besides, citizen science and community-based monitoring programs as well as remote sensing approaches (Chandler *et al.*, 2017) could also be very useful to get a reliable picture of the biodiversity distribution in the Mediterranean basin in the near future. Certainly, the corollary is that the scientific community and the public need more entomological culture in general (Basset & Lamarre, 2019).

Acknowledgments

This study has been supported by the projects BioWeb (MINECO: CGL2011-15622-E BOS), BANDENCO (JCMM: POII11-0277-5747) and IBERARTRO (SBPLY/17/180501/000492) founded by European Regional Development Fund (ERDF) through the Consejería de Educación, Ciencia y Cultura, Junta de Comunidades de Castilla-La Mancha. D.S.-F. is funded by a postdoctoral contract from the Spanish Ministry of Science and Innovation (Ramón y Cajal program; RYC2019-027446-I). We are also indebted to all the people who helped in assembling the databases, especially Arancha Caballero, Víctor Triviño, Suraya M. Vargas, Alicia B. Gómez, Belinda Bargueño, Pedro Abellán, Daniel Bruno and Félix Picazo. We thank Lucia Maltez for her meticulous English corrections. Special thanks to Dr. Yves Basset and Dr. Simone Fattorini for their valuable suggestions.

Conflict of Interest

The authors have declared that there is no conflict of interest

Authors' contributions

JML and DSF conceived the ideas and designed the methodology. All authors provided the data included in the insect databases. JML analysed the data; JML and DSF led manuscript writing, and all authors contributed to drafts and gave final approval for publication.

References

- Abellán, P. & Svenning, J.C. (2014) Refugia within refugia—patterns in endemism and genetic divergence are linked to Late Quaternary climate stability in the Iberian Peninsula. *Biological Journal of the Linnean Society*, **113**, 13–28.
- Ball-Damerow, J.E., Brenskelle, L., Barve, N., Soltis, P.S., Sierwald, P., Bieler, R., LaFrance, R., Ariño, A.H., Guralnick & R. (2019) Research applications of primary biodiversity databases in the digital age. *PLoS ONE*, **14**, e0215794.
- Balmford, A. & Gaston, K.J. (1999) Why biodiversity surveys are good value. *Nature*, **398**, 204–205.
- Basset, Y. & Lamarre, G.P. (2019) Toward a world that values insects. *Science*, **364**, 1230–1231.
- Bélanger, J. & Pilling, D. (2019) The state of the world’s biodiversity for food and agriculture. FAO Commission on Genetic Resources for Food and Agriculture Assessments, [Online]. Available: <http://www.fao.org/3/CA3129EN/ca3129en.pdf>
- Blondel, J. & Aronson, J. (1999) Biology and Wildlife of the Mediterranean Region. Oxford University Press, Oxford, UK.
- Bolker, B.M. (2008) *Ecological models and data in R*. Princeton University Press, Oxford.
- Bruno, D., Sánchez-Fernández, D., Millán, A., Ros, R.S., Sánchez-Gomez, P. & Velasco, J. (2012) Assessing the quality and usefulness of different taxonomic groups inventories in a semiarid Mediterranean region. *Biodiversity and Conservation*, **21**, 1561–1575.
- Brown, J.H. & Lomolino, M.V. (1998) Biogeography. 2nd ed. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.
- Cardoso, P., Barton, P.S., Birkhofer, K., Chichorro, F., Deacon, C., Fartmann, T., Fukushima, C.S., Gaigher, R., Habel, J.C., Hallmann, C.A., Hill, M.J., Hochkirch, A., Kwak, M.L., Mammola, S., Noriega, J.A., Orfinger, A.B., Pedraza, F., Pryke, J.S., Roque, F.O., Settele, J., Simaika, J.P., Stork, N.E., Suhling, F., Vorster, C. & Samways, M.J. (2020) Scientists’ warning to humanity on insect extinctions. *Biological Conservation*, **242**, 108426.
- Chandler, M., See, L., Copas, K., Bonde, A.M.Z., Lopez, B.C., Danielsen, F., Legind, J.K., Masinde, S., Miller Rushing, A.J., Newman, G., Rosemartin, A. & Turak, E. (2017)

432 Contribution of citizen science towards international biodiversity monitoring.
 433 *Biological Conservation*, **213**, 280–294.

434 Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through
 435 extrapolation. *Philosophical Transactions of the Royal Society B*, **345**, 101–118.

436 Cox, C.B. & Moore, P.D. (2004) Biogeography: An Ecological and Evolutionary Approach.
 437 Blackwell, Oxford, UK.

438 Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the influence of
 439 hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73–77.

440 Fattorini S (2013) Regional Insect Inventories Require Long Time, Extensive Spatial Sampling
 441 and Good Will. *PLoS ONE*, **8**, e62118.

442 Fick, S.E. & Hijmans, R.J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces
 443 for global land areas. *International Journal of Climatology* **37**, 4302–4315.

444 Flather, C.H. (1996) Fitting species-accumulation functions and assessing regional land use
 445 impacts on avian diversity. *Journal of Biogeography*, **23**, 155–168.

446 Forister, M.L., Pelton, E.M. & Black, S.H. (2019) Declines in insect abundance and diversity:
 447 We know enough to act now. *Conservation Science and Practice*, **1**, e80.

448 García-Barros, E., Munguira, M.L., Martín, J, Romo, H., Garcia-Pereira, P. & Maravalhas, E.S.
 449 (2004) Atlas de las mariposas diurnas de la Península Ibérica e islas Baleares
 450 (Lepidoptera: Papilionoidea & Hesperioidea). *Monografías de la Sociedad*
 451 *Entomológica Aragonesa*, **11**, 1-228.

452 Harvey, J.A., Heinen, R., Armbrrecht, I., Basset, Y., Baxter-Gilbert, J.H., Bezemer, T.M.,
 453 Böhm, M., Bommarco, R., Borges, P.A.V., Cardoso, P., Clausnitzer, V., Cornelisse, T.,
 454 Crone, E.E., Dicke, M., Dijkstra, K.D.B., Dyer, L., Ellers, J., Fartmann, T., Forister,
 455 M.L., Furlong, M.J., Garcia-Aguayo, A., Gerlach, J., Gols, R., Goulson, D., Habel, J.C.,
 456 Haddad, N.M., Hallmann, C.A., Henriques, S., Herberstein, M.E., Hochkirch, A.,
 457 Hughes, A.C., Jepsen, S., Jones, T.H., Kaydan, B.M., Kleijn, D., Klein, A.M., Latty,
 458 T., Leather, S.R., Lewis, S.M., Lister, B.C., Losey, J.E., Lowe, E.C., Macadam, C.R.,
 459 Montoya-Lerma, J., Nagano, C.D., Ogan, S., Orr, M.C., Painting, C.J., Pham, T.H.,
 460 Potts, S.G., Rauf, A., Roslin, T.L., Samways, M.J., Sanchez-Bayo, F., Sar, S.A.,
 461 Schultz, C.B., Soares, A.O., Thancharoen, A., Tscharntke, T., Tylianakis, J.M.,
 462 Umbers, K.D.L., Vet, L.E.M., Visser, M.E., Vujic, A., Wagner, D.L., WallisDeVries,
 463 M.F., Westphal, C., White, T.E., Wilkins, V.L., Williams, P.H., Wyckhuys, K.A.G.,
 464 Zhu, Z.R. & de Kroon, H. (2020) International scientists formulate a roadmap for insect
 465 conservation and recovery. *Nature Ecology & Evolution*, **4**, 174–176.

466 Gaston, K.J. (2000) Global patterns in biodiversity. *Nature* **405**, 220–227.

467 Gómez, A. & Lunt, D.H. (2004) Refugia within refugia: patterns of phylogeographic
 468 concordance in the Iberian Peninsula. *Phylogeography of the Southern European*
 469 *refugia* (ed. by Weiss, S. & Ferrand, N.), pp. 155–188. Springer, The Netherlands.

470 Griffiths, G.H., Eversham, B.C. & Roy, D.B. (1999) Integrating species and habitat data for
 471 nature conservation in Great Britain: data sources and methods. *Global Ecology and*
 472 *Biogeography*, **8**, 329–345.

473 Guisan, A., Thuiller, W. & Zimmermann, N.E. (2017) Habitat suitability and distribution
 474 models, with applications in R. Cambridge University Press, Cambridge, UK.

475 Goulson, D. (2019) The insect apocalypse, and why it matters. *Current Biology*, **29**, R967–
 476 971.

477 Habel, J.C., Samways, M.J. & Schmitt, T. (2019) Mitigating the precipitous decline of
 478 terrestrial European insects: Requirements for a new strategy. *Biodiversity and*
 479 *Conservation*, **28**, 1343–1360.

480 Hallmann, C.A., Sorg, M., Jongejans, E., Siepel, H., Hofland, N., Schwan, H., Stenmans, W.,
 481 Muller, A., Sumser, H., Horren, T., Goulson, D. & de Kroon, H. (2017) More than 75
 482 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS*
 483 *ONE*, **12**, 1–21.

484 Hortal, J., de Bello, F., Diniz-Filho, J.A.F, Lewinsohn, T.M. & Lobo, J.M. & Ladle, R.J. (2015)
 485 Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of*
 486 *Ecology, Evolution, and Systematics*, **46**, 523–549.

487 Hortal, J. & Lobo J.M (2005) An ED-based protocol for the optimal sampling of biodiversity.
 488 *Biodiversity and Conservation*, **14**, 2913–2947.

489 Kujala, H., Moilanen, A. & Gordon, A. (2018) Spatial characteristics of species distributions
 490 as drivers in conservation prioritization. *Methods in Ecology and Evolution*, **9**, 1121–
 491 1132.

492 Lawton, J.H., Prendergast, J.R. & Eversham, B.C. (1994) The numbers and spatial distributions
 493 of species: analyses of British data. *Systematics and conservation evaluation* (ed. by
 494 P.L. Forey, C.J. Humphries and R.I. Vane-Wright), pp. 177– 195. Clarendon Press,
 495 Oxford, UK.

496 Legendre, P. & Legendre, L. (2012) *Numerical Ecology*. Elsevier, Amsterdam, The
 497 Netherlands.

498 Lobo, J.M. (2008) Database records as a surrogate for sampling effort provide higher species
 499 richness estimations. *Biodiversity and Conservation*, **17**, 873–881.

500 Lobo, J.M. & Martín-Piera, F. (1991) La creación de un banco de datos zoológico sobre los
501 Scarabaeidae Ibero-Baleares. *Elytron*, **5**, 31–37.

502 Lobo, J.M., Baselga, A., Hortal, J., Jiménez-Valverde, A. & Gómez, J.F. (2007) How does the
503 knowledge about the spatial distribution of Iberian dung beetle species accumulates
504 over time? *Diversity and Distributions*, **13**, 772–780.

505 Lobo, J.M., Hortal, J., Yela, J.L., Millán, A., Sánchez-Fernández, D., García-Roselló, E.,
506 González-Dacosta, J., Heine, J., González-Vilas, L. & Guisande, C. (2018) KnowBR:
507 an application to map the geographical variation of survey effort and identify well-
508 surveyed areas from biodiversity databases. *Ecological Indicators*, **91**, 241–248.

509 McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models*. Chapman & Hall, New York

510 Mair, L. & Ruete, A. (2016) Explaining spatial variation in the recording effort of citizen
511 science data across multiple taxa. *PLoS ONE*, **11**, e0147796.

512 Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015) Global priorities for an effective
513 information basis of biodiversity distributions. *Nature Communications*, **6**, 8221.

514 Millán, A., Sánchez-Fernández, D., Abellán, P., Picazo, F., Carbonell, J.A., Lobo, J.M. &
515 Ribera, I. (2014). *Atlas de los coleópteros acuáticos de España peninsular*. Ministerio
516 de Agricultura, Alimentación y Medio Ambiente. Madrid.

517 Montgomery, G.A., Belitz, M.W., Guralnick, R.P. & Tingley, M.W. (2021) Standards and best
518 practices for monitoring and benchmarking insects. *Frontiers in Ecology and*
519 *Evolution*, **8**, 579193.

520 Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B. & Kent, J. (2000)
521 Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853– 858.

522 Oliveira, U., Paglia, A.P., Brescovit, A.D., de Carvalho, C.B., Silva, D.P., Rezende, D., Leite,
523 F.S.F., Batista, J.A.N., Barbosa, J.P.P.P., Stehmann, J.R., Ascher, J., de Vasconcelos,
524 M.F., De Marco, P., Löwenberg-Neto, P., Dias, P.G., Ferro, V.G. & Santos, A. (2016)
525 The strong influence of collection bias on biodiversity knowledge shortfalls of
526 Brazilian terrestrial biodiversity. *Diversity and Distributions*, **22**, 1232–1244.

527 Payo-Payo, A. & Lobo, J.M. (2016) The unpredictable characteristics of the localities where
528 new Iberian species will be discovered. *Diversity and Distributions*, **22**, 73–82.

529 Ramos, M.A., Lobo, J.M. & Esteban, M. (2001) Ten years inventorying the Iberian fauna:
530 results and perspectives. *Biodiversity and Conservation*, **10**, 19–28.

531 Reuter, H. I., Lado, L. R., Hengl, T. & Montanarella, L. (2008) Continental-scale digital soil
532 mapping using European Soil Profile Data: Soil pH, *Hamburger Beiträge zur*
533 *Physischen Geographie und Landschaftsökologie*, **19**, 91–102.

534 Romo, H., García-Barros, E. & Lobo, J.M. (2006) Identifying recorder-induced geographic
535 bias in an Iberian butterfly database. *Ecography*, **29**, 873-885.

536 Rosso, A., Aragón, P., Acevedo, F., Doadrio, I., García-Barros, E., Lobo, J.M., Munguira,
537 M.L., Monserrat, V.J., Palomo, J., Pleguezuelos, J.M., Romo, H., Triviño, V. &
538 Sánchez-Fernández, D. (2018) Effectiveness of the Natura 2000 network in protecting
539 Iberian endemic fauna. *Animal Conservation*, **21**, 262–271.

540 Ruffo, S. & Stoch, F. (2006) Checklist and distribution of the Italian fauna: 10,000 terrestrial
541 and inland water species. *Memorie del Museo Civico di Storia Naturale di Verona*, **2**.
542 Sezione Scienze della Vita, 17.

543 Sánchez-Fernández, D., Fox, R., Dennis, R.H.L. & Lobo, J.M. (2021) How complete are insect
544 inventories? An assessment of the British butterfly database highlighting the influence
545 of distribution shifts due to climatic changes. *Biodiversity and Conservation*, in press.

546 Sánchez-Fernández, D., Lobo, J.M., Abellán, P., Ribera, I. & Millán, A. (2008) Bias in
547 freshwater biodiversity sampling: the case of Iberian water beetles. *Diversity and*
548 *Distributions*, **14**, 754–762.

549 Sastre, P. & Lobo, J.M. (2009) Taxonomist survey biases and the unveiling of biodiversity
550 patterns. *Biological Conservation*, **142**, 462-467.

551 Saunders, M.E., Janes, J.K. & O’Hanlon, J.C. (2020) Moving On from the insect apocalypse
552 narrative: engaging with evidence-based insect conservation. *BioScience*, **70**, 80–89.

553 Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial
554 trends. *Frontiers in Zoology*, **4**, 11.

555 Soberón, J. & Llorente, J. (1993) The use of species accumulation functions for the prediction
556 of species richness. *Conservation Biology*, **7**, 480-488.

557 Speed, J.D.M., Bendiksby, M., Finstad, A.G., Hassel, K., Kolstad, A.L. & Prestø, T. (2018)
558 Contrasting spatial, temporal and environmental patterns in observation and specimen
559 based species occurrence data. *PLoS ONE*, **13**, e0196417.

560 Stork, N.E. (2018) How many species of insects and other terrestrial arthropods are there on
561 Earth? *Annual Review of Entomology*, **63**, 31.

562 Tiago, P., Ceia-Hasse, A., Marques, T.A., Capinha, C. & Pereira, H.M. (2017) Spatial
563 distribution of citizen science casuistic observations for different taxonomic
564 groups. *Scientific Reports*, **7**, 12832.

565 Tulloch, A.I.T., Mustin, K., Possingham, H.P., Szabo, J.K. & Wilson, K.A. (2013) To boldly
566 go where no volunteer has gone before: predicting volunteer activity to prioritize
567 surveys at the landscape scale. *Diversity and Distributions*, **19**, 465–480.

568 Wagner, D.L. (2020) Insect Declines in the Anthropocene. *Annual Review of Entomology*, **65**,
569 457–80.

570 Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R. & Stopak. D. (2021) Insect
571 decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National*
572 *Academy of Sciences of the United States of America*, 118, e2023989118.

573 Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E.M. & Willis, K.J. (2005)
574 Conservation Biogeography: assessment and prospect. *Diversity and Distributions*, **11**,
575 3–23.

576

Table 1.- Number of database records (R), species (S) and percentages of the cells with a resolution of 25 minutes, belonging to six different ordinal categories according to completeness values. Cells without any observation were categorized as 1, while cells with very few records so that their completeness cannot be calculated are categorized as 2. The remaining categories are established according to the progressively increasing completeness (C) values, from 1% to 100%.

	R	S	1	2	1<C<50	50<C<75	75<C<90	C>90
Caddisflies	13,961	444	44.76	24.08	15.86	11.33	3.68	0.28
Moths	168,507	848	24.08	11.90	30.88	19.26	10.76	3.12
Dung beetles	68,710	242	22.10	9.35	24.36	26.91	14.45	2.83
Aquatic beetles	63,498	493	18.41	10.76	33.71	27.76	8.78	0.57
Butterflies	423,820	230	9.63	1.42	10.20	18.70	32.86	27.20
Total	738,496	2,257						

Table 2.- Results of the Generalized Linear Models estimating the individual explanatory capacity of each predictor on the presence-absence (1/0) of reliable inventories in 25 minute square cells (binomial), and on the continuous completeness values in the cells with reliable inventories (continuous). %Dev is the percentage of change in deviance in a null model with no predictors. %Dev in brackets indicates the explained deviance considering only environmental variables as predictors (i.e., without the distance to the main centres of research). The values for each predictor are the parameter estimates and Wald statistic values, and their associated probabilities are indicated in brackets. Only those with an associated probability ≤ 0.05 are included. E = Elevation; GRAS = grasslands; AMT = annual mean temperature; DIST = distance to main centres of research; MTWM = maximum temperature of the warmest month; AP = annual precipitation; DMP = driest month precipitation.

	Caddisflies	Butterflies	Dung beetles	Aquatic beetles	Moths	
Binomial	%Dev	24.41 (22.57)	10.13 (9.94)	8.23 (5.13)	12.21 (11.73)	13.96 (13.53)
	E	—	—	—	1.21 (4.59; 0.03)	1.83 (5.89; 0.02)
	pH	0.91 (11.40; <0.001)	0.60 (8.32; 0.004)	—	0.74 (12.51; <0.001)	—
	GRAS	—	—	3.45 (4.57; 0.03)	—	—
	MTWM	-1.11 (4.74; 0.03)	-0.94 (4.09; 0.04)	—	—	—
	DIST	-0.52 (7.69; 0.005)	—	-0.85 (12.74; <0.001)	—	—
	AP	0.40 (4.30; 0.04)	—	—	—	—
Continuous	%Dev	22.08 (17.07)	13.79 (7.33)	22.18 (14.75)	6.55 (4.92)	19.02 (15.91)
	E	—	—	0.30 (10.34; 0.001)	—	0.13 (5.72; 0.02)
	pH	—	—	-0.12 (11.18; 0.001)	—	—
	AMT	—	—	0.35 (5.45; 0.02)	—	0.19 (4.66; 0.03)
	DIST	-0.17 (7.48; 0.006)	-0.16 (17.69; <0.001)	-0.18 (21.53; <0.001)	-0.07 (4.80; 0.03)	-0.06 (12.60; <0.001)
	MTWM	—	—	—	—	-0.17 (12.22; <0.001)
	AP	—	—	—	—	-0.14 (23.16; <0.001)
	DMP	—	—	0.11 (4.69; 0.03)	—	—

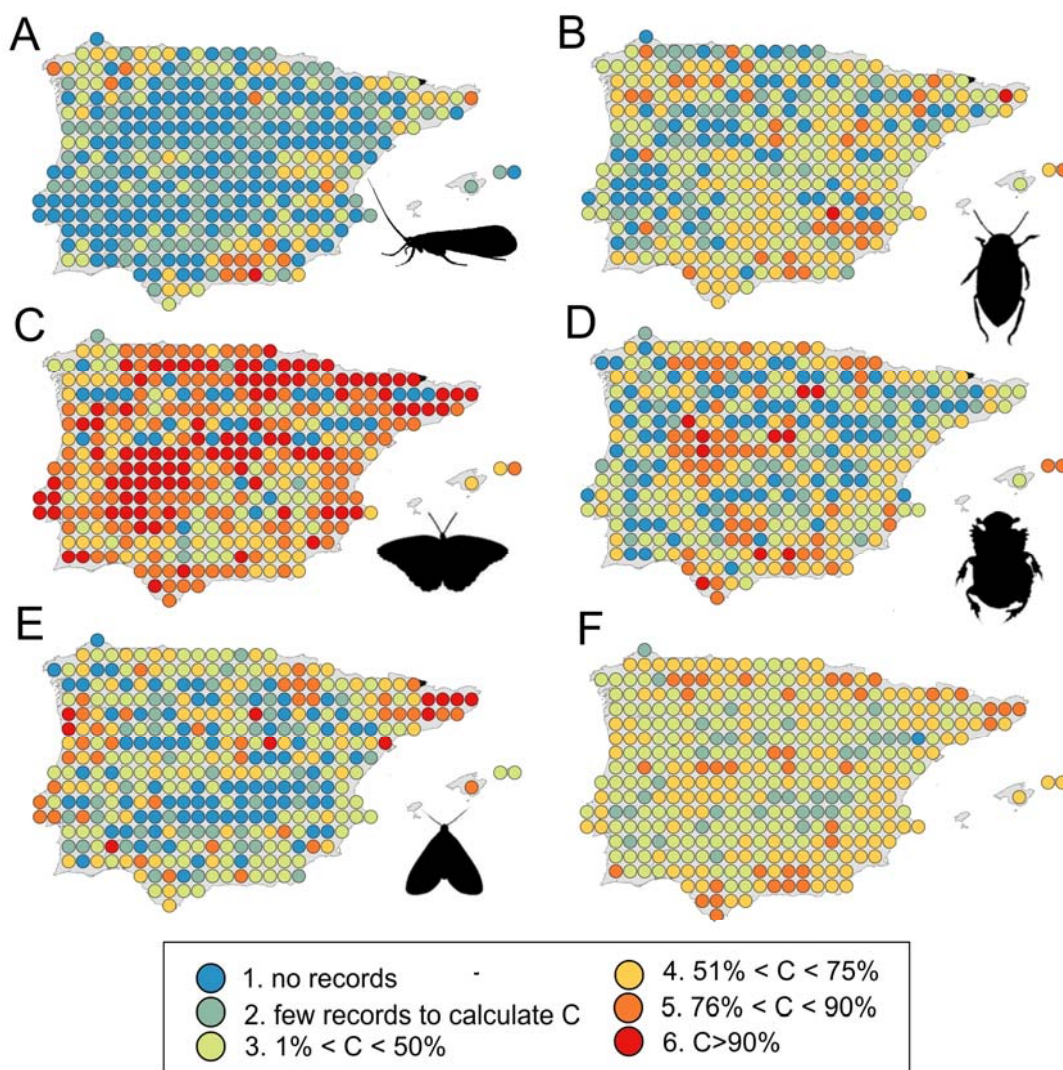


Figure 1. Categories related to the values of completeness (C) for different insect databases (A: Caddisflies; B: Aquatic beetles; C: Butterflies; D: Dung beetles; E: Moths; F: averaged values) at a spatial resolution of 25 minute cells.

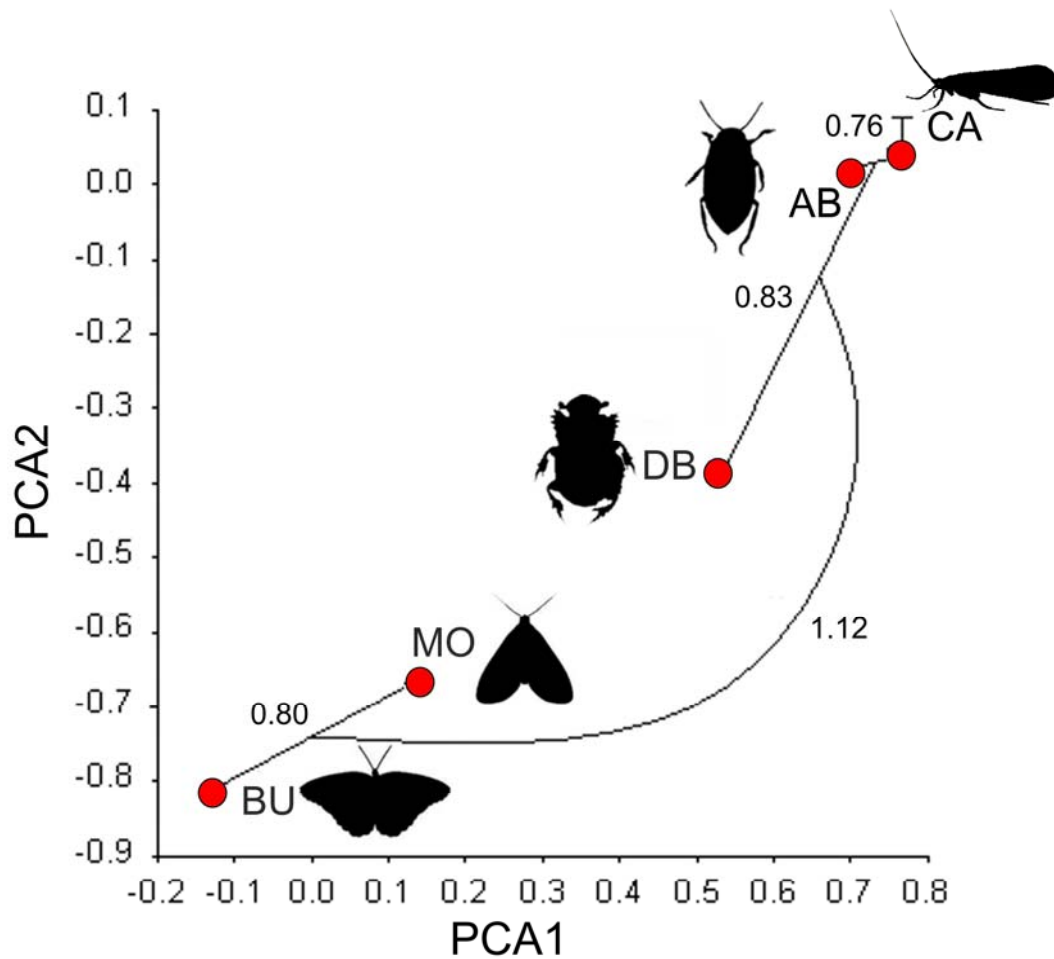


Figure 2. Position of the five insect groups in the space delimited by the factor loadings of the two main axes of a Principal Component Analysis of completeness values for all 25 minute cells of the Ibero-Balearic area. MO = Moths; BU = Butterflies; DB = Dung beetles; AB = Aquatic beetles; CA = Caddisflies. Lines connecting the five groups represent the grouping results of a Cluster Analysis based on the completeness values of these taxa. Numbers are the linkage distance among the different clusters.

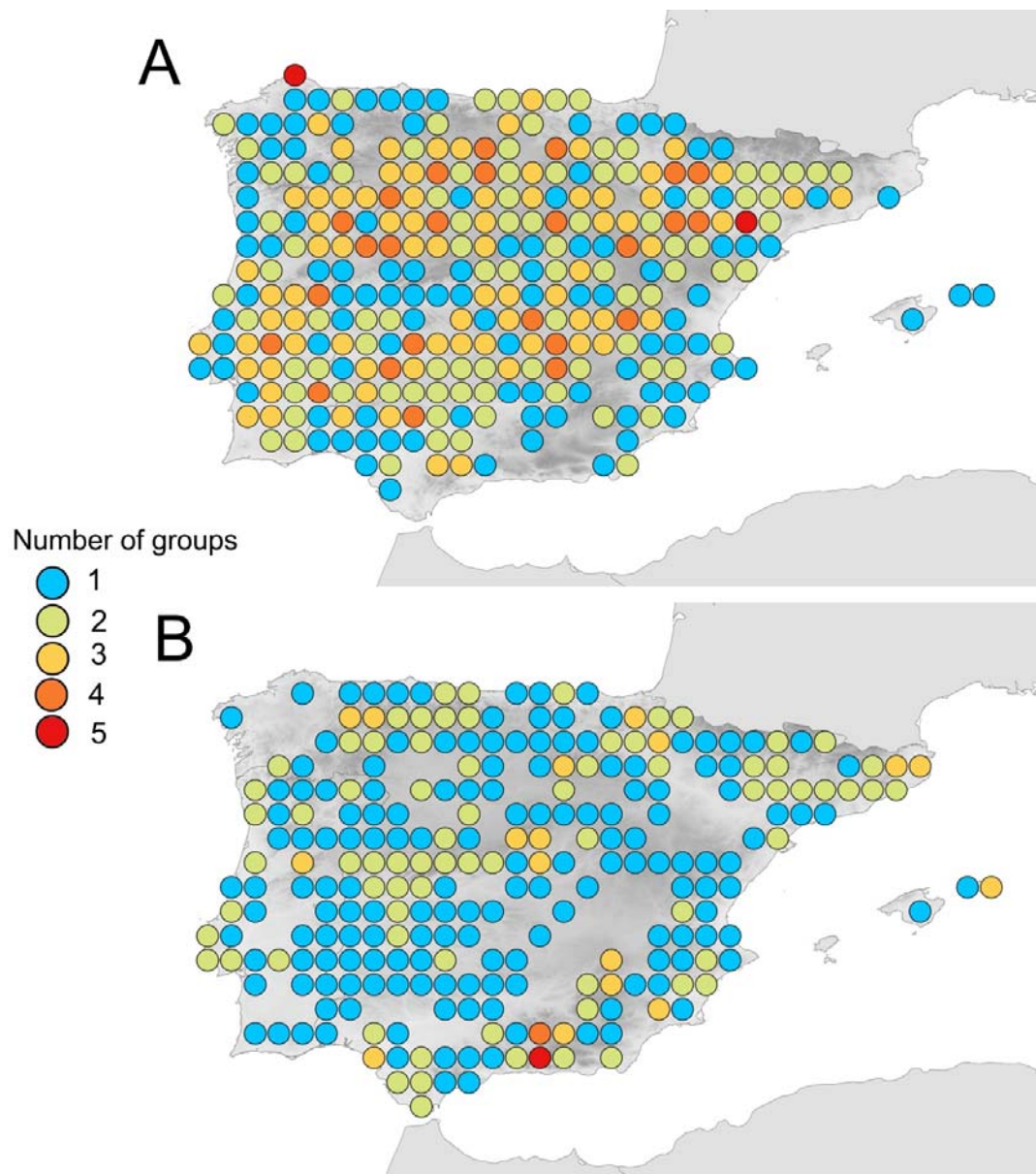


Figure 3. Number of insect groups by which each cells is identified as A) badly surveyed cells (categories 1 and 2; see Table 1) and B) relatively and well-surveyed cells (categories 5 and 6; see Table 1).