Received: 26 July 2022 | Revised: 12 December 2022 | Accepted: 14 March 2023
DOI: 10.1111/jbi.14614

RESEARCH ARTICLE

Journal of Biogeography
WILEY

European bee diversity: Taxonomic and phylogenetic patterns

Nicolas Leclercq[1] | Leon Marshall[1] | Geoffrey Caruso[2] | Kerry Schiel[2] | Timothy Weekers[1] | Luísa G. Carvalheiro[3,4] | Holger H. Dathe[5] | Michael Kuhlmann[6] | Denis Michez[7] | Simon G. Potts[8] | Pierre Rasmont[7] | Stuart P. M. Roberts[1] | Guy Smagghe[9] | Peter Vandamme[10] | Nicolas J. Vereecken[1]

[1]Agroecology Lab, Université libre de Bruxelles (ULB), Boulevard du Triomphe CP 264/02, B-1050, Brussels, Belgium

[2]Department of Geography and Spatial Planning, University of Luxembourg and Luxembourg Institute of Socio-Economic Research, Maison des Sciences Humaines, Campus Belval 11, porte des Sciences, L-4366, Esch-sur-Alzette, Luxembourg

[3]Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Brazil

[4]Center of Ecology, Evolution and Environmental Changes, Universidade de Lisboa, Lisbon, Portugal

[5]Senckenberg Deutsches Entomologisches Institut, Eberswalder Straße 90, 15374, Müncheberg, Germany

[6]Zoological Museum of Kiel University, Hegewischstr. 3, D-24105, Kiel, Germany

[7]Laboratory of Zoology, Université de Mons, Place du parc 20, 7000, Mons, Belgium

[8]Centre for Agri-Environmental Research, School of Agriculture, Policy and Development, University of Reading, Reading, RG6 6AR, UK

[9]Laboratory of Agrozoology, Department of Plant and Crops, Faculty of Bioscience Engineering, Ghent University, Coupure Links, 653, Ghent, Belgium

[10]Department of biochemistry and microbiology, Ghent University, Ghent, Belgium

Correspondence
Nicolas Leclercq, Agroecology Lab, Campus de la Plaine (CP 264/2), Université libre de Bruxelles (ULB), 1050 Brussels (Belgium).
Email: nicolas.leclercq@ulb.be

Funding information
FNRS/FWO joint programme "EOS–Excellence of Science", Grant/Award Number: 30947854

Handling Editor: Werner Ulrich

Abstract

Aim: Wild bees still face striking shortfalls in knowledge of biodiversity in key regions of the world. This includes Europe, where despite a long tradition of data gathering, the continental scale distribution patterns of wild bees have not been systematically analysed to date. This study aims to characterise large-scale biodiversity patterns to: (i) understand spatial–temporal heterogeneity in large-scale databases, (ii) locate genuine diversity hotspots and their relationship with biogeographical patterns or habitats of interests and (iii) identify understudied species and areas to further design conservation actions for most at risk species in key regions.

Location: Europe.

Taxon: Bees.

Methods: We present a continental and standardised study of bee taxonomic and phylogenetic diversity patterns in Europe, using a large compilation of occurrence records of nearly three million validated occurrence records for 1515 wild bee species.

Results: Southern and eastern Europe suffer from the largest gaps in data availability while northern and western regions benefit from better historical coverage. Our models show that higher wild bee diversity in Europe is hosted in xeric, warm areas, as highlighted by a clear latitudinal gradient. However, phylogenetic diversity is predicted to be more homogenous across Europe than taxonomic diversity, suggesting that policies and strategies targeted to protect species richness may differ from those targeting greater phylogenetic diversity.

Main conclusions: This study represents a significant advance in the characterisation of wild bee distribution patterns across Europe and is an important stepping stone towards the design of more targeted survey efforts and conservation actions of this key group of pollinators. This, in turn, will provide the data necessary to improve the spatiotemporal coverage in a context of ongoing and future Europe-wide monitoring schemes, to ultimately develop cost-effective, coordinated and evidence-based conservation actions and tailored habitat management actions that can be implemented on a smaller scale.

KEYWORDS
drivers of biodiversity change, GBIF, Hymenoptera, latitudinal gradient

1 | INTRODUCTION

Wild bees provide important ecosystem services along with playing a key role as pollinators to both wild plants and crops, yet they are currently facing multiple anthropogenic threats worldwide (Klein et al., 2018; Potts et al., 2016; Wagner, 2020). A key challenge for their conservation lies in assessing how individual bee species and communities as a whole respond to increasing and interacting drivers of decline. The availability of reliable and robust data is an essential prerequisite to ensure high quality research and evidence-based conservation actions (IPBES, 2016), yet wild bee occurrence data are typically characterised by significant taxonomic and geographical biases (Hortal et al., 2015; Orr et al., 2021; Wetzel et al., 2018). This phenomenon occurs at continental scales and below, owing to the sheer species diversity of this pollinator group, contrasting with the general lack of resources, experts, digitisation efforts in historical collections and financial support to facilitate their identification down to the species-level and databasing. As a result, the biodiversity patterns of only a limited number of popular and/or well-defined taxonomic species groups have been investigated so far (e.g. bumblebees, Rasmont et al., 2015), which hinders a deeper understanding of bee diversity sensu *lato* and at large spatial scales (see Ascher et al., 2020).

Although Europe is arguably one of the most intensively surveyed regions in the world in terms of wild bee diversity, more than 55% of all its known species of bees were described as ‘Data Deficient’ in the first IUCN Red List (Nieto et al., 2014). This first large-scale evaluation of the conservation status of European bees called for an urgent need to fill all major data gaps through coordinated long-term monitoring programmes involving both expert researchers and citizen scientists, along with gaining a better understanding and characterisation of the community structure and population trends, species ecological requirements, and modelling of wild bee species distribution (Nieto et al., 2014; Potts et al., 2020).

Statistical models can be used in conjunction with monitoring to estimate knowledge gaps, to highlight priority conservation areas (e.g. hotspots of diversity) and under-sampled regions, as well as to investigate the major drivers of large-scale diversity patterns (Hortal et al., 2015; Proença et al., 2017). Recently, Orr et al. (2021) produced the first-ever model of global bee species relative richness highlighting a bimodal latitudinal gradient followed by bee diversity and driven primarily by xeric, warm-temperate-climate with high solar radiation (Bystriakova et al., 2018; Michener, 1979; Petanidou et al., 1995; Wcislo, 1987). Likewise, taxonomic bee diversity at the European scale is expected to increase with decreasing latitude, with species richness reaching its peak in the Mediterranean region (Michener, 1979). Yet, the study by Orr et al. (2021) was performed using only publicly available records and checklists, and produced patterns of bee richness that were not always consistent with known patterns for many regions (they built their European model based on New World's model), suggesting that enhanced and improved datasets are still required to address the role of multiple drivers in structuring the diversity of wild bees at continental scales such as in Europe. Further, the distribution of phylogenetic diversity for bees remains vastly understudied, despite its importance as an alternative diversity facet (De Palma et al., 2017; Vereecken et al., 2021).

In the present study, we aim to fill some of the gaps mentioned above by investigating the distribution patterns of both taxonomic and phylogenetic diversities of 1515 wild bee species in Europe and by making available a new dataset consisting of a compilation of 1,520,434 published plus 1,584,917 restricted access (i.e. unpublished regional/national/international sources) occurrence records. Specifically, we first aim to describe patterns governing the diversity of wild bees and to characterise hotspots of bee taxonomic and phylogenetic diversities through distribution models. We then test the long-standing hypothesis that bee diversity varies across European biogeographical regions, with some regions such as the Mediterranean standing out as hotspots of diversity (Michener, 1979; Orr et al., 2021). We then highlight regions within Europe currently characterised by significant levels of deficits in data availability, based on known and predicted diversity, and that should be targeted by ongoing and future Europe-wide pollinator monitoring schemes. Finally, we quantify, compare and contrast published and access restricted databases in terms of taxonomic and geographic information to assess the inherent quality of these datasets and their suitability for large-scale biodiversity assessments.

2 | MATERIALS AND METHODS

2.1 | Dataset compilation, curation and verification

We compiled bee occurrence records for Europe from both public and restricted sources. First, the restricted dataset (hereafter referred to as STEP+) included occurrence records from five projects/experts: (i) the 'Status and Trends of European Pollinators' project (STEP, www.STEP-project.net, representing 57.3% of the STEP+ dataset) for which the data are hosted at the University of Mons (Belgium) and served as the backbone of the European Red List of Bees (Nieto et al., 2014; Potts et al., 2011); (ii) the 'Bees, Wasps & Ants Recording Society' programme for the United Kingdom data (BWARS, www.bwars.com, 36.0%); (iii) the 'National Biodiversity Data Centre' programme for Ireland data (NBDC, www.biodiversityireland.ie, 4.4%); (iv) H. Dathe's database for *Hylaeus* spp. (1.9%); and (v) M. Kuhlmann's database for *Colletes* spp. (0.4%).

Second, the public dataset (hereafter referred to as GBIF) was downloaded from Global Biodiversity Information Facility (GBIF) (https://doi.org/10.15468/dl.ssbtj7). The STEP+ and GBIF datasets were then combined into one unique dataset (hereafter referred to as All records), following curation steps and the suppression of overlapping records (Table 1). Overlap was considered for duplicated occurrences with the same geographic coordinates (rounded to four decimal degrees) and collected within the same year.

For each of the two datasets (GBIF and STEP+), only records with both coordinates and year of collection were kept. Records not identified down to the species level were removed (subspecies and variety were reduced to species level) (Table 1). Each species name was checked for validity, and eventually corrected for spelling mistakes and synonyms using (i) Michener (2007), (ii) the IUCN Red List checklist (Nieto et al., 2014) and its addition (Rasmont et al., 2017), (iii) the DiscoverLife's checklist (Ascher & Pickering, 2022), (iv) the Atlas Hymenoptera website (Atlas Hymenoptera, 2021), (v) the 'Palearctic Osmiine bees' checklist (Müller, 2019), (vi) the 'Integrated Taxonomic Information System' (ITIS, 2022) and the most up-to-date publications (e.g. Aubert, 2020; Bossert et al., 2019, 2022; Schwarz et al., 2019; Schwarz & Smit, 2020). The list of all species, their phylogenetic classification, their synonym(s) and their number of records in the 'All records' dataset is given in Table S1. The presence in Europe of species not recorded in the IUCN checklist and its update were verified and either kept or removed when recorded outside the study area (extinct species were also removed) (Nieto et al., 2014; Rasmont et al., 2017). We compiled a European checklist (Table S1) using these publications and compared it with our occurrence records to classify 'missing species', those without distribution data.

Europe was defined as all countries within the geographical bounds of the European Union due to data availability which consisted of European Union member countries plus Albania, Bosnia and Herzegovina, Kosovo, Macedonia, Montenegro, Norway, Serbia, Switzerland and the UK (Cyprus and Malta were

TABLE 1 Data filtering and number of modifications made, number of grid cells, records or species at different selection steps for each dataset of European wild bee occurrences. The numbers in brackets are the proportion (%) of the number in question after selection compared to the original number. Species richness is the number of species after all steps of data filtering for each dataset. Spatial distribution of selected cells based on standard error and well-sampled grid cells are shown in Figure 1.

Type of modification/selection	GBIF	STEP+	All records
Number of species names modifications			
Spelling	0	23	/
Synonyms	96	149	/
Number of grid cells, total in Europe = 8983 (100)			
Cells with data	6016 (67.0)	8755 (97.5)	8809 (98.1)
Selection based on standard error (Figures 1a–c, 2a and 3)	5870 (65.3)	8510 (94.7)	8571 (95.4)
Well-sampled grid cells (WSGC) (Figures 1d–f and 2b)	2048 (22.8)	2964 (33.0)	3847 (42.8)
Number of records			
Original dataset	1,520,434 (100)	1,584,917 (100)	2,974,081 (100)
Records with year	1,502,491 (98.8)	1,584,902 (100)	/
Records of species within the boundaries of the mask of Europe	1,501,924 (98.8)	1,545,921 (97.5)	/
Records after SE selection	1,497,741 (98.5)	1,536,362 (96.9)	2,965,504 (99.7)
Records in WSGC	1,457,655 (95.9)	1,313,665 (82.9)	2,782,173 (93.5)
Number of species			
Species richness	1049 (100)	1478 (100)	1515 (100)
Species richness after SE selection	1028 (98.0)	1455 (98.4)	1492 (98.5)
Species richness in WSGC	909 (86.7)	1363 (92.2)	1415 (93.4)

not included). We created a new grid of 25 km x 25 km squares using the extent of Europe to overcome two problems: some coordinates were in fact (i) the centroid of previously gridded data (maximum gridded data of ±18 km) or (ii) the coordinates of the nearest city/town from the actual sampling location. This resulted in a geographical mask covering Europe, with 8983 grid cells of 25 km × 25 km (Figure S1).

2.2 | Models of taxonomic and phylogenetic diversity

2.2.1 | Well-sampled grid cells and measures of corrected diversity

To deal with unequal sampling intensity between grid cells, we defined a set of cells for which we know the input information is very robust. For each dataset, our 'well-sampled' grid cells (WSGC) are defined as cells: (i) with a ratio of records/number of species of at least 2, (ii) with 10 or more species and (iii) that had been sampled in at least three different years. We performed a sensitivity analysis using different thresholds, which is presented in Figure S2. These selections of WSGC and their associated diversity indices were used as the response variables of the linear mixed effects models outline below (other cells were not included in the models). The final number of WSGC and the associated final species richness are described in Table 1 for each dataset.

As proxies for taxonomic diversity (TD) and phylogenetic diversity (PD), we estimated taxonomic and phylogenetic Hill Shannon diversity corrected for sampling completeness (i.e. coverage, total probability of occurrence of species observed in the cell, see Roswell et al., 2021) for all WSGC (Chao et al., 2014). Correction by coverage was chosen over size-based correction following Chao and Jost (2012). Hill diversity numbers describe species diversity by taking into account the relative abundance of species within cells (Hill, 1973). Each taxonomic Hill diversity number could be interpreted as species counts but Hill Shannon diversity ($q = 1$) will have lower values than species richness (SR, i.e. species counts, $q = 0$) when species within a community (i.e. grid cells) will have uneven distribution of abundances, which is the case for our data (Chao et al., 2014). Hill Shannon diversity emphasises neither rare nor common species (unlike SR, which is more sensitive to rare species) and represents a suitable choice when characterising ecological gradients (Roswell et al., 2021). To correct for coverage, we rarefied/extrapolated diversity at the midpoint coverage of WSGC ranges using bootstrapping (50 replications). The midpoint target was 79.76% for GBIF, 79.81% for STEP+ and 80.18% for the 'All records' dataset. We computed the corrected taxonomic and phylogenetic Hill Shannon diversity by midpoint coverage using the *estimate3D* function of the 'iNEXT.3D' package (version 0.0.1) (Chao et al., 2021). We visualised the relationship between TD and PD for the selection of WSGC to understand how these two metrics co-vary.

For PD, we started by computing a phylogenetic tree based on the hierarchical Linnean taxonomic classification (superfamily/position/family/tribe/genus/subgenus/species) (Table S1) (Danforth et al., 2006) using the 'ape' package (version 5.3) (Paradis & Schliep, 2018). Taxonomic classifications were derived and checked from a number of sources (Ascher & Pickering, 2022; Aubert, 2020; Bossert et al., 2019, 2022; Michener, 2007; Müller, 2019; Nieto et al., 2014; Rasmont et al., 2017). This tree has then been used along with the community matrix as an input in the *estimate3D* function to compute PD.

To characterise the gaps in occurrence records (see Section 2.3), it was important to compute corrected diversities for all cells with records. However, without any selection steps, some cells, when corrected, would result in large overestimations. Therefore, we only characterised gaps on selected cells that had a standard error (SE) lower than the maximum standard error from the set of WSGC. To have a baseline measure of diversity for this larger set of cells (greater than WSGC), we then computed diversity corrected at the same coverage (midpoint).

2.2.2 | Explanatory variables and diversity patterns analyses

To predict the patterns of bee diversity for each dataset, 48 relevant variables were selected as potential drivers of bee diversity. It comprised 36 climatic and geographic variables, latitude and longitude of the centroids of each cell, nine land cover (LC) variables (proportion of each land cover category in grid cells) and the sampling effort (total number of records in each cell). For all 36 environmental variables, the mean (ME) and standard deviation (SD) were calculated for each grid cell, resulting in 72 environmental variables for a total of 84 explanatory variables (see Supplementary Notes and Table S2 for rationale for selection).

For each diversity metric (taxonomic and phylogenetic diversity) computed for the WSGC (other cells were not included in the response variable), we fitted a generalised linear model (GLM), using a gamma distribution with scaled explanatory variables (see Supplementary Notes for model selection). For each final model of diversity, we used the *predict* function of the 'stats' package (version 3.6.1) (R Core Team, 2022) to predict the values of the diversity metrics for all grid cells in Europe (8983 cells in total). The predictive power of the models was estimated with the Cragg-Uhler pseudo R^2 that ranges from 0 to 1 (Nakagawa & Schielzeth, 2013). Estimates, 95% confidence intervals, and pseudo R^2 were extracted using the *export_summs* function of the 'jtools' package (version 2.1.4) (Long, 2020). To measure the performance of the final model, we performed five-fold cross-validations. We created five spatial groups of WSGC ('kmeans' separation based on decimal degrees coordinates) (Figure S3), used it each time as the testing data (estimating the prediction error) and all other cells (four other groups) as the training set (used to train and build the model). We analysed the difference between predicted and observed values for each testing set by computing the root mean squared error (RMSE) and the mean absolute error (MAE) and using the 'caret' package (version 6.0-92) (Kuhn, 2022).

2.3 | Predicting priority needs for under-sampled areas

To identify the European regions with low sampling effort using the 'All records' dataset, we used the observed taxonomic diversity for cells with records which were selected based on their standard error (8571 cells, see Section 2.2.1 and Table 1) and the predicted taxonomic diversity for the same cells. We calculated the proportion of observed diversity given the predicted diversity. This allowed us to characterise differences between these two metrics and was used as a proxy for the sampling completeness of every grid cell. Indeed, when the predicted diversity was higher than the raw diversity for the same cell (proportion < 100%), we could expect this cell to require a more thorough sampling/research focus, with parts of that diversity still to be described. On the other hand, when the raw diversity was higher than the prediction (proportion > 100%), it could mean that the cell had already been thoroughly sampled. To discuss the patterns of these results on a larger scale than the cell level, we eventually averaged the explained proportions at the country level. For visual clarity and to facilitate discussion of the results, we divided these proportions into six categories: (i) highly under-sampled (0%–33%), (ii) under-sampled (34%–66%), (iii) moderately under-sampled (67%–99%), (iv) moderately well-sampled (100%–133%), (v) well-sampled (134%–166%) and (vi) very well-sampled (>166%).

2.4 | Patterns of diversity within biogeographical regions

To explore if wild bee diversity in Europe is explained by variation in biogeographical regions, we considered here (i) the observed diversity as the diversity values computed for the selection of WSGC for the 'All Records' dataset (3847 cells) and (ii) the predicted diversity as the predicted values for all cells (8983 cells) resulting from the model ('All records'). We first intersected each cell with the biogeographical regions layer of Europe (EEA, 2016) (Figure S1). When cells were overlapping multiple regions, the one with the largest area covering the cell was retained. We then performed four pairwise non-parametric Wilcoxon tests with adjusted *p*-values using the Benjamini & Hochberg method of the 'stats' package (version 3.6.1) (R Core Team, 2022) to test if biogeographical regions presented significant difference in terms of observed or predicted diversities (two tests for each type of diversity: the taxonomic and phylogenetic diversity).

To explore the extent to which biogeographical regions host dissimilar assemblages of wild bees, we performed beta-diversity analyses within each biogeographical region using the presence/absence community matrix for all cells with records (8809 cells). For this, we computed beta-diversity (Sørensen dissimilarity index, βsør) values for each pair of cells using the 'betapart' package (version 1.5.1) (Baselga et al., 2018). We analysed the variation of the total dissimilarity (βsør) within each biogeographical region and its two components, that is, the turnover (βsim) and the nestedness (βsne). The aim was to characterise if a specific region was facing more dissimilarity within its 'own' cells than another region within 'theirs', and if this dissimilarity was more driven either by the turnover (species replacement) or by the nestedness (loss/gain of species) component (Baselga, 2010).

All statistical analyses were performed in RStudio (RStudio Team, 2022) for R version 4.1.3 for Windows (R Core Team, 2022).

3 | RESULTS

3.1 | European bee diversity

3.1.1 | Model of taxonomic bee diversity

The published (GBIF) and restricted access (STEP+) datasets and models showed large differences (Figure 1), and their comparison is described in the Supplementary Notes. In brief, The STEP+ dataset covered significantly more of Europe than the GBIF dataset at the levels of cells with data (97.5% vs. 67.0%), of WSGC (33.0% vs. 22.8%), and had also an important higher number of species (1478 vs. 1049) (Table 1). As expected, the lower quality data of GBIF resulted in an over-prediction of presences for a small subset of grid cells, for example, in Spain and Switzerland (Figure 1g) and did not result in a clear gradient of diversity for cells further north, as they presented similar ranges of predicted values. The STEP+ model presented more convincing patterns (Figure 1h). A low number of cells presented suspected over-predictions as well, especially in Switzerland, but more restricted than for GBIF.

The combination of these two datasets resulted in the 'All records' dataset with 2,974,081 records. It corresponded to 1515 species with records, 1012 being shared between GBIF and STEP+, 466 species only detected in STEP+ and 37 species only in GBIF (Table 1, Table S1 and Figures S4 and S5). Before any selection of well-sampled grid cells (WSGC), 403 species were already missing in the dataset, having no occurrence records, as our mask of Europe has 1918 species according to the checklists (Table S1; Figures S4 and S5). Besides the fact that some species are completely missing from the dataset, there were also 659 species with less than 50 occurrences, a remarkably low abundance for a dataset of this size (Table S1).

The 'All records' dataset covered 98.1% of all cells in Europe before any selection of cells (Table 1; Figure S6). However, the raw values of taxonomic diversity (after selection based on standard error) indicated massive unequal samplings with some cells showing high diversity values and most of the cells with extremely low values driven by low number of records and species collected (mean = 27.85 ± 23.85, min = 1.00, max = 190.09) (Figure 1c). The

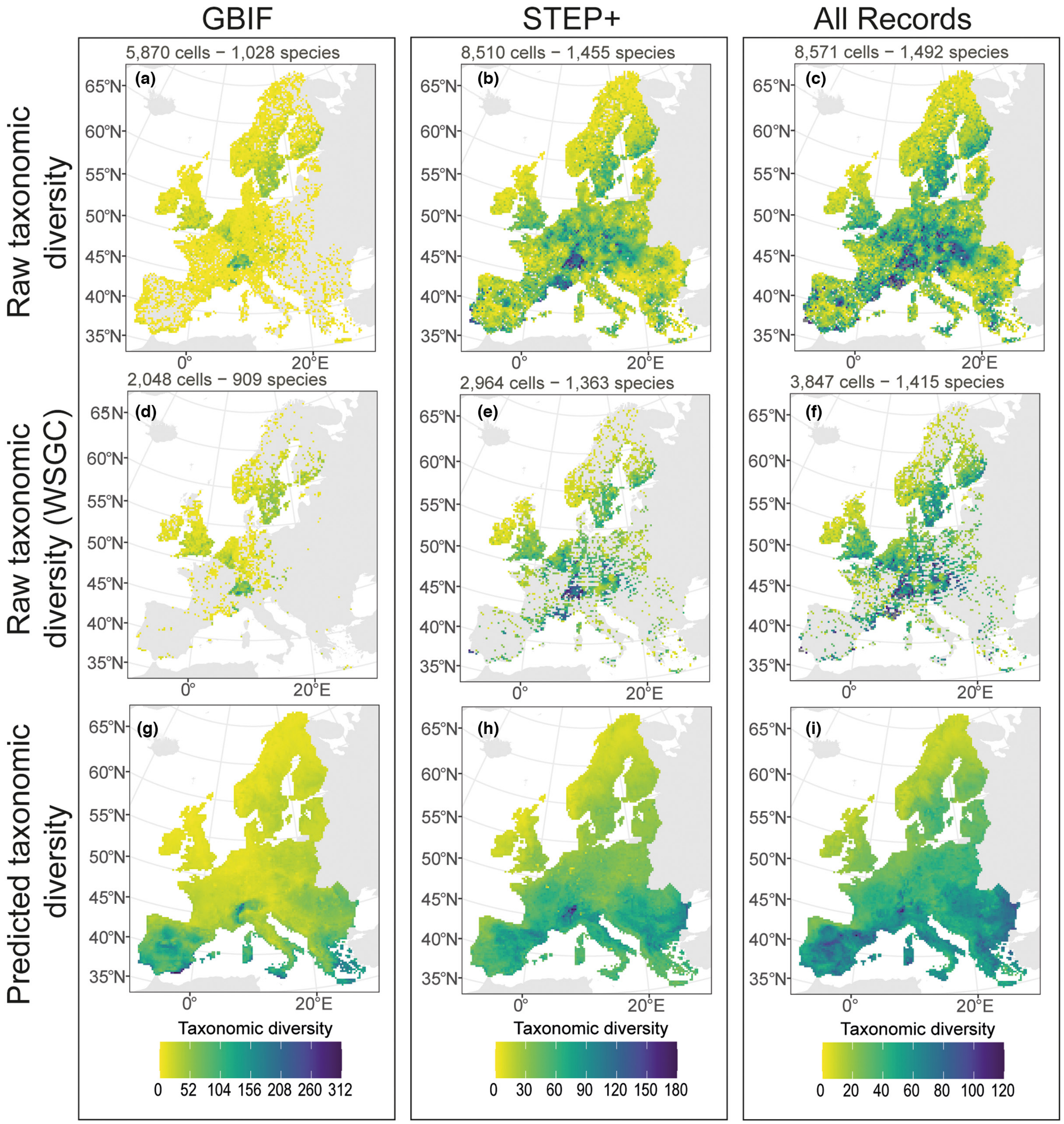

FIGURE 1 Raw and predicted values of European wild bee taxonomic diversity for GBIF, STEP+ and 'All records' datasets: (i) the distribution of raw values of taxonomic diversity (after selection based on the standard error) for (a) GBIF, (d) STEP+ and (g) 'All records' datasets; (ii) the selection of the well-sampled grid cells (WSGC) for (b) GBIF, (e) STEP+ and (h) 'All records' datasets; and (iii) the predicted values for all cells (8983 cells) of taxonomic diversity for (c) GBIF, (f) STEP+ and (i) 'All records' datasets. For each dataset, the displayed gradient of taxonomic diversity is based on the range of the predicted values (minimum and maximum values). Cells with no records are shown in grey. Number of cells and number of species are displayed for the raw diversities. European map projection: EPSG:3035, ETRS89-extended/LAEA Europe.

selection of WSGC resulted in 3847 cells (42.8% of all cells), with most records (93.5%, 2,782,173 records; mean = 723.21 ± 2415.38 records, min = 20, max = 43,056 records) and number of species (93.4%, 1415 species; mean = 65.93 ± 57.40 species, min = 10, max = 504 species) being kept during this process (Table 1; Figure 1f; Figure S6). It meant that the 'All records' dataset also presented massive gaps of WSGC in Europe, but it covered better than GBIF or STEP+ (Table 1; Table S3; Figure 1d–f). These selected cells tend

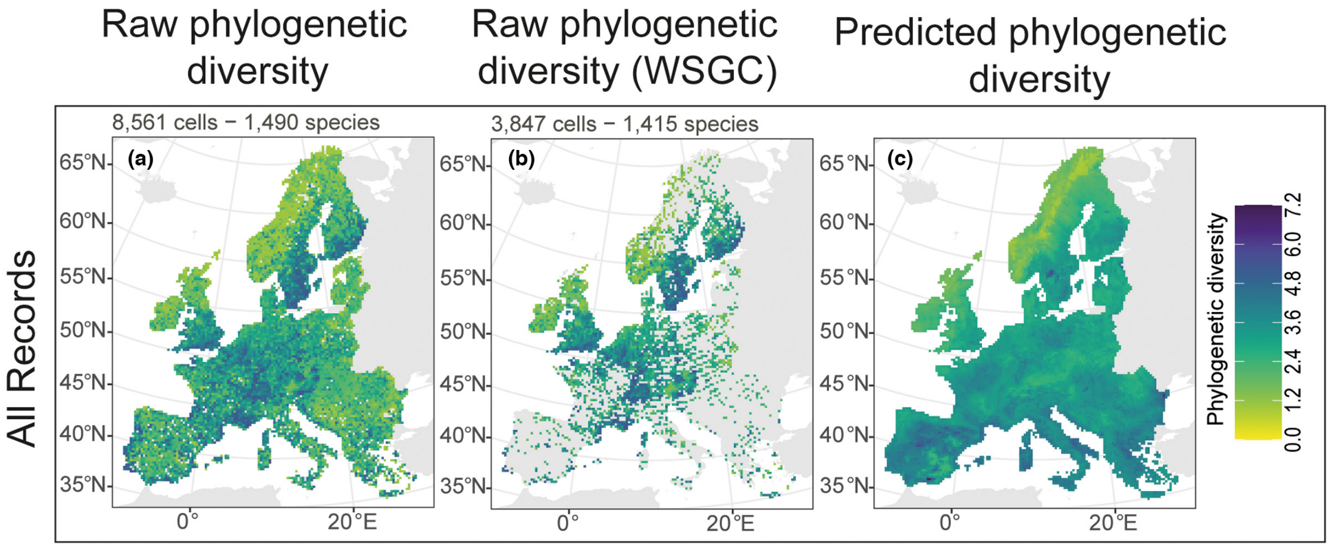

FIGURE 2 Raw and predicted values of phylogenetic European wild bee diversity for the 'All records' dataset: (a) the distribution of raw values of phylogenetic diversity (after selection based on the standard error), (b) the selection of well-sampled grid cells (WSGC) and (c) the predicted values for all cells (8983 cells) of phylogenetic diversity. The displayed gradient of phylogenetic diversity is based on the range of the predicted values. Cells with no records are shown in grey. Number of cells and number of species are displayed for the raw diversities. European map projection: EPSG:3035, ETRS89-extended/LAEA Europe.

to better cover northern regions (mean latitude = 52.85 and mean longitude = 9.97), compared to southern and eastern countries (with less than 20% of cover for most concerned countries) (Table S3). The selection of WSGC allowed for the removal of highly under-sampled cells (mean = 32.88 ± 24.98, min = 1.03, max = 190.09) (Figure 1f; Figure S6).

The resulting model of taxonomic bee diversity predicted the most convincing patterns among the three datasets with less outliers and smoother gradients (pseudo-$R^2 = 0.42$) (predicted taxonomic diversity: mean = 40.01 ± 19.34, min = 6.28, max = 118.87) (Figure 1i). The five-fold cross validations showed that our model was performant to predict on the testing set (mean MAE = 17.60 and max observed TD = 190.09, resulting in an error of 9.26%). The two most diverse groups (groups 4 and 5) showed the highest errors between predicted and observed values (Figure S3). Overall, taxonomic bee diversity showed a noticeable latitudinal gradient with higher diversity expected in southern Europe (see Table S3). Indeed, the southern countries had the highest mean predicted diversities, but overall, there was much lower disparity between the countries than for the other datasets. For instance, Macedonia, the country with the highest average predicted diversity per cell (mean = 67.54 ± 10.28, min = 48.07, max = 90.87), was less than five times higher than the country with lowest predicted diversity, that is Norway (mean = 13.41 ± 5.27, min = 6.34, max = 33.32) (Table S3). Outliers were again observed in the Alpine region near Switzerland (mean = 53.22 ± 14.44, min = 35.46, max = 112.37).

The taxonomic model using the 'All records' dataset included 11 covariates and six quadratic terms (pseudo-$R^2 = 0.42$) (Table S4). All estimates and confidence intervals are shown in Table S4. Of the set of selected variables used in this model, solar radiation (Srad_ME, estimate = 0.36, CI = [0.31, 0.40], $p < 0.001$) and then the aridity (AI_ME, −0.30, [−0.35, −0.25], $p < 0.001$) showed the highest estimates to predict taxonomic bee diversity (Table S4). Therefore, high solar radiation (available energy) (maximum at 17,900 kJ m^{-2} day^{-1}) and more arid climates resulted in high taxonomic bee diversity.

3.1.2 | Model of phylogenetic bee diversity

Using again the 'All records' dataset, we computed the phylogenetic diversity for all selected cells based on standard error (Table 1). Overall, those cells presented moderate and homogeneous values of phylogenetic diversity (mean = 2.73 ± 1.12, min = 1.00, max = 5.89), except for some areas that showed low diversities mainly in the Atlantic and Alpine parts of Scandinavia, in the North of the British Isles or in parts of eastern Europe (Figure 2a).

The same set of WSGC as the one used for the taxonomic model was used here (3487 WSGC). The selection improved the mean raw diversity (mean = 3.02 ± 1.11, min = 1.00, max = 5.89) (Figure 2b). In WSGC, the taxonomic and phylogenetic diversities were strongly correlated. However, PD tended to quickly increase even for low TD while, at the opposite edge of the spectrum, PD reached a plateau at the highest TD values (Figure S7).

The resulting phylogenetic model (pseudo-$R^2 = 0.34$) and its predicted values presented fewer clear patterns than the taxonomic diversity, with lower impact of the latitudinal gradient due to these homogenous values (mean = 3.16 ± 0.77, min = 0.78, max = 7.07) (Figure 2c). At the country level, all countries presented homogeneous mean predicted diversity with low disparity even between the country with the highest mean predicted diversity, that is Macedonia (4.01 ± 0.31, min = 3.43, max = 4.50), and the country with the lowest, that is Norway (1.73 ± 0.55, min = 0.78, max = 3.47). Overall, the

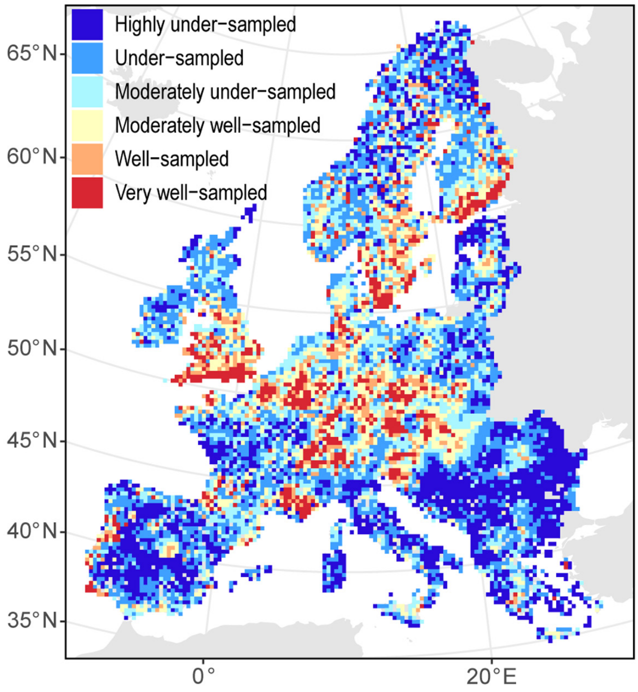

FIGURE 3 Proportion of observed wild bee taxonomic diversity (after selection based on the standard error) given the predicted taxonomic diversity for 8571 cells using the 'All records' dataset. For visual clarity, six categories were created. Proportions were categorised such as: (i) highly under-sampled (0%–33%), (ii) under-sampled (34%–66%), (iii) moderately under-sampled (67%–99%), (iv) moderately well-sampled (100%–133%), (v) well-sampled (134%–166%) and (vi) very well-sampled (>166%). Above 100% proportions imply that the observed diversity was higher than the predicted diversity. It means that it was comparatively more well-sampled than cells with similar conditions. European map projection: EPSG:3035, ETRS89-extended/LAEA Europe.

southern countries showed the highest mean predicted phylogenetic diversity (Table S3).

The phylogenetic model included 11 covariates and five quadratic terms (pseudo-$R^2 = 0.34$). Overall, variables presented similar estimates to predict phylogenetic diversity, but the maximum temperature of warmest month (Bio5_ME, estimate = 0.12, CI = [0.10, 0.14], $p < 0.001$) had the highest estimate (Table S4).

3.2 | Predicting priority needs for under-sampled areas

In this section, we used the set of 8571 cells that were selected based on their standard error and for which we have computed raw taxonomic diversities that we compared with predicted values. Overall, the results showed that 72.3% of Europe lacked occurrence records (see Hortal et al., 2015; Wetzel et al., 2018), as 25.7% of the cells were considered as 'highly under-sampled' (proportion of observed diversity given the predicted diversity between 0% and 33%), 28.0% as 'under-sampled' (34%–66%), and 18.6% as 'moderately under-sampled' (67%–99%) (Figure 3). At the country level, eastern countries, then southern countries showed the lowest mean proportions (Table S3). On the other hand, four countries showed the highest mean proportions which were considered as 'well-sampled' and 'very well-sampled' (>133%, Belgium, the Czech Republic, Luxembourg and Switzerland) (Table S3).

3.3 | Patterns of diversity within biogeographical regions

Here, we checked for the variation of observed diversity values in the 3847 WSGC and predicted diversity values (8983 cells) for taxonomic and phylogenetic diversity within each biogeographical region. The observed values of both taxonomic and phylogenetic diversities presented higher overlaps between biogeographical regions than the predicted diversity (Figure 4; Table S5). Indeed, at the taxonomic level, all pairs of regions were significantly different from each other in their predicted diversity except for the pair Black Sea-Steppe (Table S5; Figure 4a). For the phylogenetic diversity, all pairs were significantly different in their predicted phylogenetic diversity (Table S5; Figure 4b). Overall, the Steppe (mean of predicted taxonomic diversity = 77.99 ± 6.32, min = 69.87, max = 95.91; and mean of predicted phylogenetic diversity = 4.12 ± 0.47, min = 3.27, max = 6.39), the Black Sea (77.20 ± 9.74, min = 56.14, max = 93.18; 4.33 ± 0.45, min = 3.82, max = 6.03) and the Mediterranean (59.70 ± 12.15, min = 24.08, max = 118.87; 3.90 ± 0.49, min = 2.23, max = 7.07) regions, followed by the Pannonian region (58.58 ± 5.16, min = 46.23, max = 77.04; 3.56 ± 0.22, min = 3.09, max = 4.71), were predicted as hotspots of bee diversity, with significant differences compared to the other regions (Table S5; Figure 4a,b). The Mediterranean region presented the highest predicted value for a cell but a high variability due to its large size (1902 cells for predicted diversity, 366 WSGC) in comparison with the restricted ranges for the Steppe (59 cells, 1 WSGC) and the Black Sea (25 cells, 4 WSGC) regions (Table S5). Moreover, two groups of values are distinguishable in the predicted diversity of the Alpine region (Figure 4a,b). The group with the lowest values corresponds to the Alpine region located in Scandinavia that receive low solar radiation compared to the other group located at lower latitudes (Figure 4c).

Using the 'All records' dataset and its presence/absence community matrix for all cells with records (8809 cells), we tested how varied the dissimilarity in species composition within each region. In Figure 4d, we presented the variability of the total dissimilarity (βsør) and its two components, that is, turnover (βsim) and nestedness (βsne), within each region. Overall, the dissimilarity of each region was driven by turnover rather than nestedness (means and standard deviations for each region and for each component are displayed in Table S5). This was particularly evident for the Mediterranean region, which presented the highest total dissimilarity (mean βsør = 0.89 ± 0.10), mainly explained by the large species turnover (mean βsim = 0.77 ± 0.21) between the Mediterranean cells. On the other hand, the dissimilarities within the three other regions predicted as the most diverse, that is the Black Sea (mean

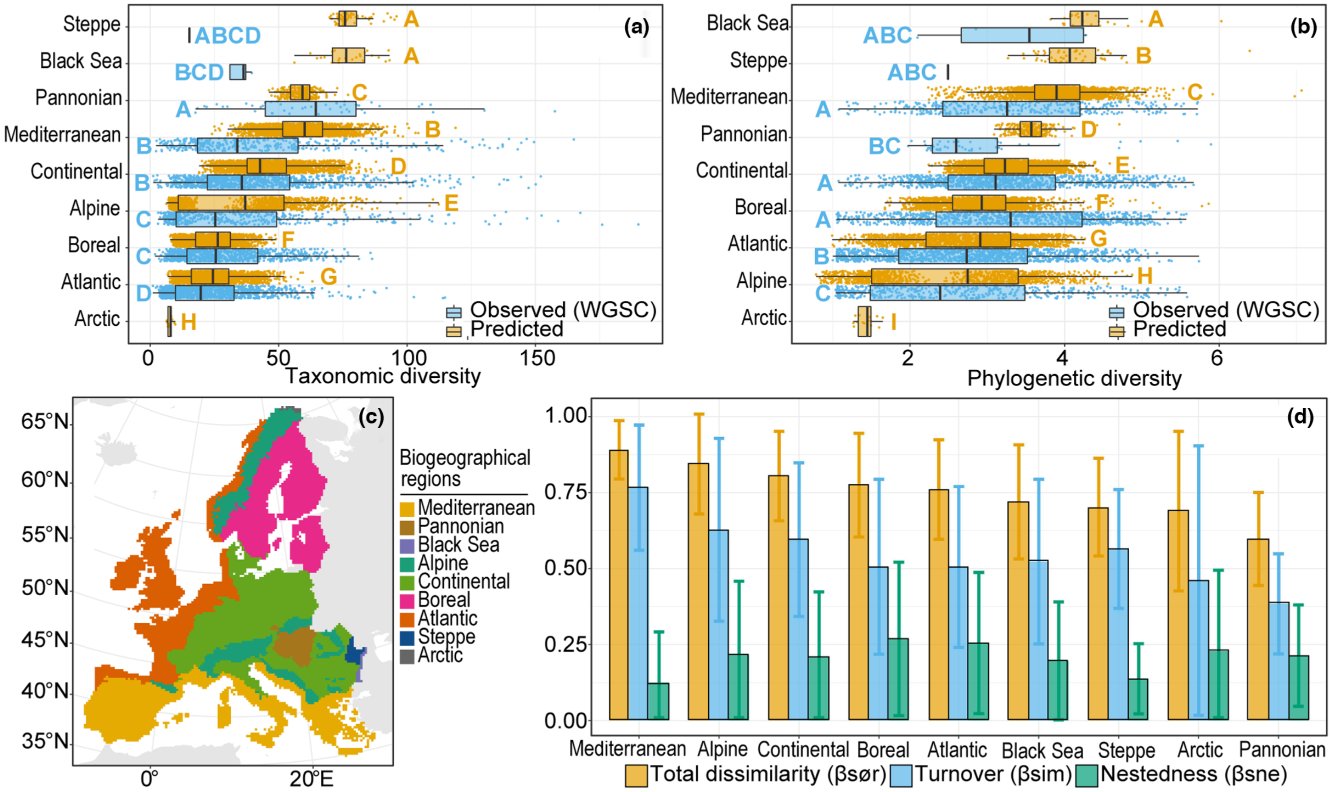

FIGURE 4 Patterns for the 'All records' dataset of observed European wild bee diversity in well-sampled grid cells (WSGC) and predicted diversity (8983 cells) for both (a) taxonomic diversity and (b) phylogenetic diversity and (d) dissimilarity within each biogeographical region (c). (a and b) The dots indicate values of either the observed diversities (blue) or the predicted diversities (orange). Boxes indicate the inter-quartile ranges (IQR), that is, Q3–Q1, where Q1 and Q3 are the first and third quartiles, respectively, and the vertical lines indicate the medians (Q2). Potential outliers are dots that are outside the range of [Q1 − 1.5 × IQR, Q3 + 1.5 × IQR], that is, outside the range of the horizontal lines. Coloured letters indicate significant different groups of regions using the pairwise non-parametric Wilcoxon test for both observed and predicted set of values (Table S5). (d) Mean and standard deviation of total dissimilarity (βsør) and its two components: turnover (βsim) and nestedness (βnes) for each biogeographical region. European map projection: EPSG:3035, ETRS89-extended/LAEA Europe.

βsør = 0.70 ± 0.16), Steppe (0.72 ± 0.19) and Pannonian (0.60 ± 0.15) regions, were comparatively lower (Table S5).

4 | DISCUSSION

Our study provides a comprehensive modelling of European biodiversity patterns for all bees using occurrence records sourced from both public and restricted datasets. Here, we show that higher wild bee diversity in Europe is hosted in (semi-)arid and warm regions that receive high solar radiation. Even though our models highlight an absence of high-quality data in many regions, they confirm the relevance of widespread occurrence data to determine hotspots of diversity, and call for increased attention, monitoring, and conservation efforts in understudied regions.

We show that any attempt to model the diversity of bees in Europe with publicly available data only (GBIF) is largely irrelevant to understand biogeographic patterns (see Boyd et al., 2022; Orr et al., 2021). By contrast, the larger, private dataset (STEP+) is much more informative in terms of species richness and spatial coverage. However, these data are from different large-scale scientific projects, combining data from standardised sampling projects or museum data, and access to those depends on the goodwill of the data owners. Here, we make this aggregated data available to unlock the possibility to reproduce our analyses with further improved data. Indeed, even though the 'All records' dataset that we used in this study is the most complete database of cleaned and verified biological records to date, there is still a wealth of existing but unpublished data relevant to Europe within the scientific community and networks of naturalists, both within and outside Europe (Wetzel et al., 2018). Therefore, we stress the importance of establishing practical tools to structure and standardise data sharing (EU Pollinators Initiative, 2021; IPBES, 2016; Potts et al., 2011) and to ensure the consistency of records including validated species names; information on the sex of individuals whenever possible; accurate and detailed geolocation of specimens; date of collection; collection method used; and specimen identification information (including name of the person in charge of the identification).

Furthermore, we show that we mainly lack data from southern and eastern Europe (Wetzel et al., 2018). Indeed, northern and western regions are historically well studied, because of a long history of amateur naturalists and more recently well-funded scientific projects compared to the rest of Europe (Nieto et al., 2014; Wetzel et al., 2018). A global effort from the scientific community and naturalists to focus on under-sampled regions is necessary for conservation at the continental scale. For example, ongoing projects in Spain and Portugal are addressing this (Sánchez-Fernández et al., 2022; Wood, 2022; Wood et al., 2021, 2022). Our study helps to inform further survey schemes that are being implemented by the European Pollinator Monitoring Scheme (EU PoMS) in several countries. More specifically, we provided evidence for the need of improving the schemes and policy actions by integrating targeted surveys and conservation efforts in these key regions (Potts et al., 2020).

In this study, we provide a checklist for all European species up to 2020 following most recent publications (Ascher & Pickering, 2022; Atlas Hymenoptera, 2021; Aubert, 2020; Bossert et al., 2019, 2022; ITIS, 2022; Michener, 2007; Müller, 2019; Nieto et al., 2014; Rasmont et al., 2017; Schwarz et al., 2019; Schwarz & Smit, 2020). This list highlights a range of understudied species that the scientific community should prioritise in future sampling programmes. Long-term conservation of wild bee diversity in Europe requires knowledge of the present distribution and an understanding of temporal dynamics in the face of increasing anthropogenic factors.

Our results focused on taxonomic and phylogenetic diversity; the next logical step will be to model the functional diversity. Previous studies have explored, at relatively large-scale and for subsets of well-studied species, the relations between functional diversity and landscape/land-use gradients or habitats and have highlighted the importance of using alternative diversity indices (Carrié et al., 2017; De Palma et al., 2015, 2017; Normandin et al., 2017). However, broad patterns of the diversity of functional traits at the European scale are not known and many gaps remain to cover all species in Europe with every trait described. We need a compilation of all these traits into a single and publicly available database. This is the aim of the work of Roberts, S.P.M., with the help of several organisations to build an open access European wild bee trait dataset. This database was created and updated as part of the ALARM and STEP projects and will be used within the framework of the ORBIT project (ORBIT, 2022; Potts et al., 2011; Settele et al., 2005). We hope it will unlock the possibility of exploring the evolutionary structure and functioning of biodiversity. Integrating functional perspectives may offer another framework for predicting the species responses to climate change and increasing anthropogenetic pressures, and characterising functional diversity as a proxy for pollination services which would be a determinant indicator for policy (De Palma et al., 2017; Potts et al., 2020; Vereecken et al., 2021).

Our standardised data cleaning and filtering processes enabled the characterisation of European diversity patterns using existing occurrence data while previous measures were either based on countries checklists, performed at regional levels or for some bee groups (Bystriakova et al., 2018; Michener, 1979; Petanidou et al., 1995; Wcislo, 1987). We see these models and this study as an important first step towards the prediction of European bee diversity. However, we showed that strong data improvements are needed, especially in Southern and Eastern Europe, to improve drastically the performance of such models, species distribution modelling being one of the major objectives in order to characterise diversity hotspots and conservation areas. Predicted latitudinal gradient demonstrates that the Black Sea, Mediterranean, Pannonian and Steppe biogeographical regions are home to a unique and high wild bee diversity in Europe. They are indeed unique regions characterised by their own distinctive weather and plant communities (Médail, 2008; Rundel & Cowling, 2013) that may coincide with the ecological optima of many wild bees. Furthermore, the Mediterranean region is comparatively much larger than the Steppe or Black Sea regions, and it definitely drives the fact that it arrives only third in most predicted diverse regions due to a higher variability of predicted diversity. Additionally, the Mediterranean region was the region with the highest absolute predicted diversity for a cell and the highest within dissimilarity, driven by a remarkably high turnover of species in comparison with other regions. We argue that more abundant quality data are needed in these regions to refine these results.

Of the variables considered in this study, climatic variables were the most important to predict bee diversity (Bystriakova et al., 2018; Kammerer et al., 2021; Michener, 1979; Orr et al., 2021; Petanidou et al., 1995; Wcislo, 1987), although it is important to note that climatic variables were predominant in our variable list. Taxonomic diversity was driven by high solar radiation (energy) in xeric areas (aridity). Phylogenetic diversity was also found optimal in arid climates with high temperatures during the warmest month. These results all confirmed and provided quantified evidence for Michener's (1979) pioneering description of bee biogeography and supported Orr et al.'s (2021) findings, namely that European diversity reflects a latitudinal gradient and bee diversity is higher in xeric and warm temperate regions of the world.

Phylogenetic diversity showed more homogeneity in observed and predicted values across regions than taxonomic diversity, reinforcing the importance of alternative diversity metrics, as they may capture distinct perspectives of diversity (De Palma et al., 2017; Vereecken et al., 2021). The relationship between both metrics is the main reason behind this result, i.e. PD tended to increase rapidly even for low TD and a plateau was reached at the highest TD values. This might be explained by the redundancy among species' contributions to the phylogenetic diversity as opposed to taxonomic diversity and by the evolutionary convergence among phylogenetically distant species. Moreover, we can hypothesise that phylogenetic diversity is more affected by local factors than by large-scale factors, such as the variability of flora and nesting opportunities (Carvalheiro et al., 2021), the surrounding landscape (Kammerer et al., 2016; Le Féon et al., 2010), the local intensification (Deguines et al., 2014), the local climate conditions or the abundance of managed species (Weekers et al., 2022). To confirm that different scales might highlight different drivers (Hurlbert & Jetz, 2007; Rahbek, 2004; Willig et al., 2003), it would be interesting to compare the dependence of each diversity metric on variables ranging from local to large geographic scale. Apart from the scale aspect, the choice of different resolutions might also be of importance and should be tested in further studies. Recent studies also suggest that drivers such as air ozone pollution, pests and diseases, pesticides, extreme weather events, soil type, competition, plants and pollinators networks may be valuable to include in future exercises (Phillips et al., 2018; Rollin et al., 2022; Settele et al., 2016), but are currently unavailable for large-scale analyses.

Present wild bee distributions depend on intercontinental and other natural barriers, both historical and present climatic and vegetational conditions, and the dispersal ability of each species to reach suitable habitats (Michener, 2007). To disentangle these factors and to characterise wild bee distribution patterns at species level, it is imperative to fill the data gaps highlighted in the present study. We argue that wild bee knowledge is hindered by a historical bias towards well-studied regions with a long tradition of natural history collecting. Our study highlights under-studied species and regions and can be used to select areas for monitoring. Our standardised method with data cleaning, filtering, and modelling can be replicated in different regions, at different scales or resolutions, and can easily be updated and maintained alongside future data improvements from key regions and species. Our study helps to inform further research and monitoring schemes and policy actions that need to be integrated through the EU PoMS (Potts et al., 2020). Moreover, our call for improved data on key species in key regions is an important deliverable for initiatives such as the EU Pollinators Initiative (EPI) that aims to address wild pollinators declines and inform an updated European Red List (Nieto et al., 2014; Potts et al., 2020).

ACKNOWLEDGEMENTS

The FNRS/FWO joint programme 'EOS—Excellence of Science' for the project 'CliPS: Climate change and its effects on Pollination Services (project 30947854)' and the 'Fonds Van Buuren & Fondation Jaumotte Demoulin' funded this study. We thank all people who collected and maintained the databases used in this study and that were obtained from: (i) the European Commission Framework Programme (FP) 7 via the Status and Trends of European Pollinators (STEP) collaborative project (grant no. 244090, www.STEP-project.net); (ii) the Bees, Wasps and Ants Recording Society (BWARS, www.bwars.com); (iii) the Banque de Données Fauniques Gembloux-Mons (BDFGM, www.atlashymenoptera.net); (iv) the National Biodiversity Data Centre (NBDC, www.biodiversityireland.ie); and (v) the Global Biodiversity Information Facility (GBIF, www.gbif.org). No permits were needed for this study.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available on dryad, https://doi.org/10.5061/dryad.5tb2rbp8n

ORCID

Nicolas Leclercq https://orcid.org/0000-0002-3317-6622

REFERENCES

Ascher, J. S., Marshall, L., Meiners, J., Yanega, D., & Vereecken, N. J. (2020). Heterogeneity in large-scale databases and the role of climate change as a driver of bumble bee decline. *Science*, *367*(6478), 685–688.

Ascher, J. S., & Pickering, J. (2022). *Discover life bee species guide and world checklist* (Hymenoptera: Apoidea: Anthophila). http://www.discoverlife.org/mp/20q?guide=Apoidea_species

Atlas Hymenoptera. (2021). *Atlas Hymenoptera*. http://www.atlashymenoptera.net

Aubert, M. (2020). Proposition de clé d'identification des Eucerini (Hymenoptera: Anthophila) de France continentale—Version provisoire. *Observatoire des Abeilles*, 45. https://oabeilles.net/projets/eucerini

Baselga, A. (2010). Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, *19*(1), 134–143. https://doi.org/10.1111/j.1466-8238.2009.00490.x

Baselga, A., Orme, C. D. L., Villeger, S., De Bortoli, J., & Leprieur, F. (2018). *Betapart:Partitioning Beta diversity into turnover and Nestedness components*. https://cran.r-project.org/web/packages/betapart/betapart.pdf

Bossert, S., Murray, E. A., Almeida, E. A. B., Brady, S. G., Blaimer, B. B., & Danforth, B. N. (2019). Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. *Molecular Phylogenetics and Evolution*, *130*(2), 121–131. https://doi.org/10.1016/j.ympev.2018.10.012

Bossert, S., Wood, T. J., Patiny, S., Michez, D., Almeida, E. A. B., Minckley, R. L., Packer, L., Neff, J. L., Copeland, R. S., Straka, J., Pauly, A., Griswold, T., Brady, S. G., Danforth, B. N., & Murray, E. A. (2022). Phylogeny, biogeography and diversification of the mining bee family Andrenidae. *Systematic Entomology*, *47*(2), 283–302. https://doi.org/10.1111/syen.12530

Boyd, R. J., Aizen, M. A., Barahona-Segovia, R. M., Flores-Prado, L., Fontúrbel, F. E., Francoy, T. M., Lopez-Aliste, M., Martinez, L., Morales, C. L., Ollerton, J., Pescott, O. L., Powney, G. D., Saraiva, A. M., Schmucki, R., Zattara, E. E., & Carvell, C. (2022). Inferring trends in pollinator distributions across the neotropics from publicly available data remains challenging despite mobilization efforts. *Diversity and Distributions*, *January*, 1–12. https://doi.org/10.1111/ddi.13551

Bystriakova, N., Griswold, T., Ascher, J. S., & Kuhlmann, M. (2018). Key environmental determinants of global and regional richness and endemism patterns for a wild bee subfamily. *Biodiversity and Conservation*, *27*(2), 287–309. https://doi.org/10.1007/s10531-017-1432-7

Carrié, R., Andrieu, E., Cunningham, S. A., Lentini, P. E., Loreau, M., & Ouin, A. (2017). Relationships among ecological traits of wild bee communities along gradients of habitat amount and fragmentation. *Ecography*, *40*(1), 85–97. https://doi.org/10.1111/ecog.02632

Carvalheiro, L. G., Bartomeus, I., Rollin, O., Timóteo, S., & Tinoco, C. F. (2021). The role of soils on pollination and seed dispersal. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1834), 20200171. https://doi.org/10.1098/rstb.2020.0171

Chao, A., Chiu, C.-H., & Jost, L. (2014). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 297–324. https://doi.org/10.1146/annurev-ecolsys-120213-091540

Chao, A., Henderson, P. A., Chiu, C.-H., Moyes, F., Hu, K.-H., Dornelas, M., & Magurran, A. E. (2021). Measuring temporal change in alpha diversity: A framework integrating taxonomic, phylogenetic and functional diversity and the iNEXT.3D standardization. *Methods in Ecology and Evolution*, *12*, 1–15. https://doi.org/10.1111/2041-210X.13682

Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, *93*(12), 2533–2547. https://doi.org/10.1890/11-1952.1

Danforth, B. N., Sipes, S., Fang, J., & Brady, S. G. (2006). The history of early bee diversification based on five genes plus morphology. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(41), 15118–15123. https://doi.org/10.1073/pnas.0604033103

De Palma, A., Kuhlmann, M., Bugter, R., Ferrier, S., Hoskins, A. J., Potts, S. G., Roberts, S. P. M., Schweiger, O., & Purvis, A. (2017). Dimensions of biodiversity loss: Spatial mismatch in land-use impacts on species, functional and phylogenetic diversity of European bees. *Diversity and Distributions*, *23*(12), 1435–1446. https://doi.org/10.1111/ddi.12638

De Palma, A., Kuhlmann, M., Roberts, S. P. M., Potts, S. G., Börger, L., Hudson, L. N., Lysenko, I., Newbold, T., & Purvis, A. (2015). Ecological traits affect the sensitivity of bees to land-use pressures in European agricultural landscapes. *Journal of Applied Ecology*, *52*(6), 1567–1577. https://doi.org/10.1111/1365-2664.12524

Deguines, N., Jono, C., Baude, M., Henry, M., Julliard, R., & Fontaine, C. (2014). Large-scale trade-off between agricultural intensification and crop pollination services. *Frontiers in Ecology and the Environment*, *12*(4), 212–217. https://doi.org/10.1890/130054

EEA. (2016). *Biogeographical regions*. https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3

EU Pollinators Initiative. (2021). EU Pollinators Initiative. https://ec.europa.eu/environment/nature/conservation/species/pollinators/policy_en.htm

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, *54*(2), 427–432. https://doi.org/10.2307/1934352

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Hurlbert, A. H., & Jetz, W. (2007). Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(33), 13384–13389. https://doi.org/10.1073/pnas.0704469104

IPBES. (2016). *The assessment report of the intergovernmental science-policy platform on biodiversity and ecosystem services on pollinators, pollination and food production* (S.G. Potts, V. L. Imperatriz-Fonseca, & H. T. Ngo (Eds.)). Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. https://doi.org/10.5281/zenodo.3402856

ITIS. (2022). *Integrated Taxonomic Information System*. https://www.itis.gov/

Kammerer, M. A., Biddinger, D. J., Joshi, N. K., Rajotte, E. G., & Mortensen, D. A. (2016). Modeling local spatial patterns of wild bee diversity in Pennsylvania apple orchards. *Landscape Ecology*, *31*(10), 2459–2469. https://doi.org/10.1007/s10980-016-0416-4

Kammerer, M. A., Goslee, S. C., Douglas, M. R., Tooker, J. F., & Grozinger, C. M. (2021). Wild bees as winners and losers: Relative impacts of landscape composition, quality, and climate. *Global Change Biology*, *27*(6), 1250–1265. https://doi.org/10.1111/gcb.15485

Klein, A. M., Boreux, V., Fornoff, F., Mupepele, A. C., & Pufal, G. (2018). Relevance of wild and managed bees for human well-being. *Current Opinion in Insect Science*, *26*, 82–88. https://doi.org/10.1016/j.cois.2018.02.011

Kuhn, M. (2022). *Caret: Classification and regression training*. https://cran.r-project.org/package=caret

Le Féon, V., Schermann-Legionnet, A., Delettre, Y., Aviron, S., Billeter, R., Bugter, R., Hendrickx, F., & Burel, F. (2010). Intensification of agriculture, landscape composition and wild bee communities: A large scale study in four European countries. *Agriculture, Ecosystems & Environment*, *137*(1–2), 143–150. https://doi.org/10.1016/j.agee.2010.01.015

Long, J. A. (2020). *Jtools: Analysis and presentation of social scientific data*. https://cran.r-project.org/package=jtools

Médail, F. (2008). Mediterranean. In *Encyclopedia of ecology* (Vol. *54*, pp. 2296–2308). Elsevier. https://doi.org/10.1016/B978-008045405-4.00348-7

Michener, C. D. (1979). Biogeography of the bees. *Annals of the Missouri Botanical Garden*, *66*(3), 277. https://doi.org/10.2307/2398833

Michener, C. D. (2007). *The bees of the world* (2nd ed.). The John Hopkins University Press.

Müller, A. (2019). *Palaearctic Osmiine Bees*. ETH Zürich. https://blogs.ethz.ch/osmiini/

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nieto, A., Roberts, S. P. M., Kemp, J., Rasmont, P., Kuhlmann, M., García Criado, M., Biesmeijer, J. C., Bogusch, P., Dathe, H. H., De la Rúa, P., De Meulemeester, T., Dehon, M., Dewulf, A., Ortiz-Sánchez, F. J., Lhomme, P., Pauly, A., Potts, S. G., Praz, C. Q., Window, J., & Michez, D. (2014). *European red list of bees*. Publication Office of the European Union.

Normandin, É., Vereecken, N. J., Buddle, C. M., & Fournier, V. (2017). Taxonomic and functional trait diversity of wild bees in different urban settings. *PeerJ*, *5*(3), e3051. https://doi.org/10.7717/peerj.3051

ORBIT. (2022). *ORBIT Project—Taxonomic Resources for European Bees*. https://orbit-project.eu

Orr, M. C., Hughes, A. C., Chesters, D., Pickering, J., Zhu, C.-D., & Ascher, J. S. (2021). Global patterns and drivers of bee distribution. *Current Biology*, *31*(3), 451–458. https://doi.org/10.1016/j.cub.2020.10.053

Paradis, E., & Schliep, K. (2018). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Petanidou, T., Ellis, W. N., & Ellis-Adam, A. C. (1995). Ecogeographical patterns in the incidence of brood parasitism in bees. *Biological Journal of the Linnean Society*, *55*(3), 261–272. https://doi.org/10.1006/BIJL.1995.0041

Phillips, B. B., Shaw, R. F., Holland, M. J., Fry, E. L., Bardgett, R. D., Bullock, J. M., & Osborne, J. L. (2018). Drought reduces floral resources for pollinators. *Global Change Biology*, *24*(7), 3226–3235. https://doi.org/10.1111/gcb.14130

Potts, S. G., Biesmeijer, J. C., Bommarco, R., Felicioli, A., Fischer, M., Jokinen, P., Kleijn, D., Klein, A.-M., Kunin, W. E., Neumann, P., Penev, L. D., Petanidou, T., Rasmont, P., Roberts, S. P. M., Smith, H. G., Sørensen, P. B., Steffan-Dewenter, I., Vaissière, B. E., Vilà, M., … Schweiger, O. (2011). Developing European conservation and mitigation tools for pollination services: Approaches of the STEP (status and trends of European pollinators) project. *Journal of Apicultural Research*, *50*(2), 152–164. https://doi.org/10.3896/IBRA.1.50.2.07

Potts, S. G., Dauber, J., Hochkirch, A., Oteman, B., Roy, D. B., Ahnre, K., Biesmeijer, K., Breeze, T., Carvell, C., Ferreira, C., Fitzpatrick, Ú., Isaac, N. J. B., Kuussaari, M., Ljubomirov, T., Maes, J., Ngo, H., Pardo, A., Polce, C., Quaranta, M., … Vujic, A. (2020). *Proposal for an EU pollinator monitoring scheme*. Publications Office of the European Union. https://doi.org/10.2760/881843

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., Hill, R., Settele, J., & Vanbergen, A. J. (2016). Safeguarding pollinators and their values to human well-being. *Nature*, *540*(7632), 220–229. https://doi.org/10.1038/nature20588

Proença, V., Martin, L. J., Pereira, H. M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R. D., Honrado, J. P., Jürgens, N., Opige, M., Schmeller, D. S., Tiago, P., & van Swaay, C. A. M. (2017). Global biodiversity monitoring: From data sources to essential biodiversity variables. *Biological Conservation*, *213*, 256–263. https://doi.org/10.1016/j.biocon.2016.07.014

R Core Team. (2022). *R: A language and environment for statistical computing*. https://www.r-project.org/

Rahbek, C. (2004). The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, *8*(2), 224–239. https://doi.org/10.1111/j.1461-0248.2004.00701.x

Rasmont, P., Devalez, J., Pauly, A., Michez, D., & Radchenko, V. G. (2017). Addition to the checklist of IUCN European wild bees (Hymenoptera: Apoidea). *Annales de La Société Entomologique de France (N.S.)*, *53*(1), 17–32. https://doi.org/10.1080/00379271.2017.1307696

Rasmont, P., Franzen, M., Lecocq, T., Harpke, A., Roberts, S., Biesmeijer, K., Castro, L., Cederberg, B., Dvorak, L., Fitzpatrick, U., Gonseth, Y., Haubruge, E., Mahe, G., Manino, A., Michez, D., Neumayer, J., Odegaard, F., Paukkunen, J., Pawlikowski, T., … Schweiger, O. (2015). Climatic risk and distribution atlas of European bumblebees. *BioRisk*, *10*, 1–236. https://doi.org/10.3897/biorisk.10.4749

Rollin, O., Aguirre-Gutiérrez, J., Yasrebi-de Kom, I. A. R., Garratt, M. P. D., de Groot, G. A., Kleijn, D., Potts, S. G., Scheper, J., & Carvalheiro, L. G. (2022). Effects of ozone air pollution on crop pollinators and pollination. *Global Environmental Change*, *75*(May), 102529. https://doi.org/10.1016/j.gloenvcha.2022.102529

Roswell, M., Dushoff, J., & Winfree, R. (2021). A conceptual guide to measuring species diversity. *Oikos*, *130*(3), 321–338. https://doi.org/10.1111/oik.07202

RStudio Team. (2022). *RStudio: Integrated development for R*. RStudio Inc. http://www.rstudio.com/

Rundel, P. W., & Cowling, R. M. (2013). Mediterranean-climate ecosystems. In *Encyclopedia of biodiversity* (Vol. *5*, pp. 212–222). Elsevier. https://doi.org/10.1016/B978-0-12-384719-5.00245-8

Sánchez-Fernández, D., Yela, J. L., Acosta, R., Bonada, N., García-Barros, E., Guisande, C., Heine, J., Millán, A., Munguira, M. L., Romo, H., Zamora-Muñoz, C., & Lobo, J. M. (2022). Are patterns of sampling effort and completeness of inventories congruent? A test using databases for five insect taxa in the Iberian Peninsula. *Insect Conservation and Diversity*, *15*, 406–415. https://doi.org/10.1111/icad.12566

Schwarz, M., & Smit, J. (2020). Fünf neue Nomada-Arten aus der West-Paläarktis. *Linzer Biologische Beiträge*, *39*(1), 683–694.

Schwarz, M., Smit, J., & Ockermüller, E. (2019). Weitere neue paläarktische Bienen aus der Gattung Nomada Scopoli, 1770 (Hymenoptera: Apidae). *Entomofauna*, *40*(1), 3–29.

Settele, J., Bishop, J., & Potts, S. G. (2016). Climate change impacts on pollination. *Nature Plants*, *2*(7), 16092. https://doi.org/10.1038/nplants.2016.92

Settele, J., Hammen, V., Hulme, P., Karlson, U., Klotz, S., Kotarac, M., Kunin, W., Marion, G., O'Connor, M., Petanidou, T., Peterson, K., Potts, S., Pritchard, H., Pysek, P., Rounsevell, M., Spangenberg, J., Steffan-Dewenter, I., Sykes, M., Vighi, M., … Kuhn, I. (2005). ALARM: Assessing LArge-scale environmental risks for bio- diversity with tested methods. *Gaia Ecological Perspectives for Science and Society*, *14*, 69–72.

Vereecken, N. J., Weekers, T., Leclercq, N., De Greef, S., Hainaut, H., Molenberg, J.-M., Martin, Y., Janssens, X., Noël, G., Pauly, A., Roberts, S. P. M., & Marshall, L. (2021). Insect biomass is not a consistent proxy for biodiversity metrics in wild bees. *Ecological Indicators*, *121*, 107132. https://doi.org/10.1016/j.ecolind.2020.107132

Wagner, D. L. (2020). Insect declines in the Anthropocene. *Annual Review of Entomology*, *65*(1), 457–480. https://doi.org/10.1146/annurev-ento-011019-025151

Wcislo, W. T. (1987). The roles of seasonality, host synchrony, and behavious in the evolutions and distributions of nest parasitesin Hymenoptera (Insecta), with special reference to bees (Apoidea). *Biological Reviews*, *62*(4), 515–542. https://doi.org/10.1111/j.1469-185X.1987.tb01640.x

Weekers, T., Marshall, L., Leclercq, N., Wood, T. J., Cejas, D., Drepper, B., Hutchinson, L., Michez, D., Molenberg, J.-M., Smagghe, G., Vandamme, P., & Vereecken, N. J. (2022). Dominance of honey bees is negatively associated with wild bee diversity in commercial apple orchards regardless of management practices. *Agriculture, Ecosystems & Environment*, *323*, 107697. https://doi.org/10.1016/j.agee.2021.107697

Wetzel, F. T., Bingham, H. C., Groom, Q., Haase, P., Kõljalg, U., Kuhlmann, M., Martin, C. S., Penev, L., Robertson, T., Saarenmaa, H., Schmeller, D. S., Stoll, S., Tonkin, J. D., & Häuser, C. L. (2018). Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. *Biological Conservation*, *221*, 78–85. https://doi.org/10.1016/j.biocon.2017.12.024

Willig, M. R., Kaufman, D. M., & Stevens, R. D. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, *34*(1), 273–309. https://doi.org/10.1146/annurev.ecolsys.34.012103.144032

Wood, T. J. (2022). Two new overlooked bee species from Spain (Hymenoptera: Anthophila: Andrenidae, Apidae). *Osmia*, *10*(February), 1–12. https://doi.org/10.47446/OSMIA10.1

Wood, T. J., Ghisbain, G., Michez, D., & Praz, C. J. (2021). Revisions to the faunas of Andrena of the Iberian Peninsula and Morocco with the descriptions of four new species (Hymenoptera: Andrenidae). *European Journal of Taxonomy*, *758*(758), 147–193. https://doi.org/10.5852/ejt.2021.758.1431

Wood, T. J., Molina, F. P., & Bartomeus, I. (2022). A new Andrena species (Hymenoptera: Andrenidae) from the overlooked Doñana protected areas of southern Spain. *Belgian Journal of Entomology*, *126*(January), 1–13.

BIOSKETCH

Nicolas Leclercq is interested in the patterns of biodiversity at large scales. His work often focuses on wild bee diversity. He and the other authors collaborate on questions of wild bee diversity and their conservation in their own geographical region.

Author contributions: The study was conceptualised by Nicolas Leclercq, Leon Marshall, Geoffrey Caruso, Kerry Schiel, Timothy Weekers and Nicolas J. Vereecken. Apart from them, all other authors are listed alphabetically. Holger H. Dathe, Michael Kuhlmann, Simon G. Potts, Pierre Rasmont and Stuart P. M. Roberts provided data. Nicolas Leclercq analysed the data. Nicolas Leclercq, Leon Marshall, Timothy Weekers and Nicolas J. Vereecken led the writing of the manuscript. All authors have read and agreed with to this version of the manuscript.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Leclercq, N., Marshall, L., Caruso, G., Schiel, K., Weekers, T., Carvalheiro, L. G., Dathe, H. H., Kuhlmann, M., Michez, D., Potts, S. G., Rasmont, P., Roberts, S. P. M., Smagghe, G., Vandamme, P., & Vereecken, N. J. (2023). European bee diversity: Taxonomic and phylogenetic patterns. *Journal of Biogeography*, *00*, 1–13. https://doi.org/10.1111/jbi.14614