ECOGRAPHY

Research article

Completeness analysis for over 3000 United States bee species identifies persistent data gap

Paige R. Chesshire[1], Erica E. Fischer[2], Nicolas J. Dowdy[3], Terry L. Griswold[4], Alice C. Hughes[5], Michael C. Orr[6,11], John S. Ascher[7], Laura M. Guzman[8], Keng-Lou James Hung[9], Neil S. Cobb[10] and Lindsie M. McCabe[4]

[1]Dept of Biological Sciences, Northern Arizona Univ., Flagstaff, AZ, USA
[2]Centre for the History of Science, Technology and Medicine, King's College London, London, UK
[3]Dept of Zoology, Milwaukee Public Museum, Milwaukee, WI, USA
[4]United States Dept of Agriculture, Agriculture Research Service (USDA-ARS), Logan, UT, USA
[5]School of Biological Sciences, Univ. of Hong Kong, Hong Kong, Hong Kong
[6]Key Laboratory of Zoological Systematics and Evolution, Inst. of Zoology, Chinese Academy of Sciences, Beijing, China
[7]Dept of Biological Sciences, National Univ. of Singapore, Singapore, Singapore
[8]Marine and Environmental Section, Dept of Biological Sciences, Univ. of Southern California, Los Angeles, CA, USA
[9]Oklahoma Biological Survey, Univ. of Oklahoma, Norman, OK, USA
[10]Biodiversity Outreach Network, Mesa, AZ, USA
[11]Entomologie, Staatliches Museum für Naturkunde Stuttgart, Stuttgart, Germany

Correspondence: Paige R. Chesshire (prc43@nau.edu)

Native bee species in the United States provide invaluable pollination services. Concerns about native bee declines are growing, and there are calls for a national monitoring program. Documenting species ranges at ecologically meaningful scales through coverage completeness analysis is a fundamental step to track bees from species to communities. It may take decades before all existing bee specimens are digitized, so projections are needed now to focus future research and management efforts. From 1.923 million records, we created range maps for nearly 88% (3158 species) of bee species in the contiguous United States, provided the first analysis of inventory completeness for digitized specimens of a major insect clade, and perhaps most important, estimated spatial completeness accounting for all known bee specimens in USA collections, including undigitized bee specimens. Completeness analyses were very low (3–37%) across four examined spatial resolutions when using the currently available bee specimen records. Adding a subset of observations from community science data sources did not significantly increase completeness, and adding a projected 4.7 million undigitized specimens increased completeness by only an additional 12–13%. Assessments of data, including projected specimen records, indicate persistent taxonomic and geographic deficiencies. In conjunction with expedited digitization, new inventories that integrate community science data with specimen-based documentation will be required to close these gaps. A combined effort involving both strategic inventories and accelerated digitization campaigns is needed for a more complete understanding of USA bee distributions.

Keywords: Apoidea, biodiversity, Hymenoptera, species ranges

www.ecography.org

Introduction

Evidence continues to support the decline of both animal (Ceballos et al. 2010, 2020) and plant (Gray 2019) species, documenting the initial phase of the sixth mass extinction (Leakey and Lewin 1996). Yet data are frequently inconsistent across space, time and taxa, and assessing patterns requires an understanding of what data are available and how representative they are in these domains. For most arthropod species, we cannot reliably delineate their distribution, much less understand the factors mediating realized niches. How many arthropod species will go extinct before we have a critical mass of information to inform conservation measures? There has been significant digitization coupled with community science efforts for many vertebrates, especially birds, but less so for plants, and less than a 10% digitization rate for invertebrates. It could easily take decades to transcribe all invertebrate data (Borsch et al. 2020). However, there has been enough digitization to start tracking progress and identify taxonomic and biogeographic gaps while simultaneously increasing rates of digitization.

Arthropods comprise 60% of all taxa, but most occurrence data are locked up in the one billion insect specimen labels located in the 1001 arthropod collections throughout the world (Cobb 2022). We can start targeting certain arthropod clades that have enough data to project occurrences (e.g. butterflies, dragonflies and ants) and initiate gap-filling inventories while finishing the digitization of existing specimens. Because of their functional importance and documented declines, bees are an important, signature taxon for developing more specific and strategic plans for mobilizing biodiversity data for all arthropods. First, there is a critical mass of specimen data; ca 25% of the existing bee specimens in USA collections have been transcribed, compared to a 6% average for all arthropods (Cobb et al. 2019). Second, through an extensive survey of USA collections, we estimate that roughly 4.7 million USA bee specimens have yet to be transcribed. We use 1.923 million publicly available digitized records, including both specimen and observation data for nearly 88% of bee species (3158 species) in the contiguous USA to project estimates of how digitization of the remaining ca 4.7 million specimens could enhance our knowledge of bee species distributions.

Understanding bee species distributions and the communities that they form is fundamental to developing research programs and regional-to-national inventory and/or monitoring programs aimed at pollinator conservation. Furthermore, in the era sometimes described as the 'insect apocalypse', reliable information on species ranges and population data is needed to facilitate management that safeguards these species and the crucial services they provide (Goulson 2019). An increasingly popular method for delineating insect distributions uses transcribed specimen label data with additional observational records from a variety of sources (Cobb et al. 2019, Di Cecco et al. 2021, Shirey et al. 2021). These data establish critical information on locality, date and taxonomy needed to conduct a spectrum of evolutionary-ecological research, which in turn can facilitate management and help answer large-scale environmental questions (Hampton et al. 2013, Meyer et al. 2015, Lobo et al. 2018). Occurrence records help visualize the past and current distributions of bee species on local to continental scales (Orr et al. 2021) and are the basis for tracking range shifts and linking impacts of changes in climate or land management.

Inventory completeness analyses use specimen records and/or community science data to measure the number of species documented in an area compared to those expected to be present (Soberón et al. 2007, Meyer et al. 2015, Lobo et al. 2018, Nava-Bolanos et al. 2022). Completeness analyses are useful for understanding deficiencies or strengths in biodiversity data from regional to global scales and can inform gap-filling surveys before species become a conservation concern (Pelayo-Villamil et al. 2018, La Sorte and Somveille 2020, Shirey et al. 2021). Bees comprise the monophyletic group Anthophila, encompassing more than 20 000 described species worldwide (Ascher and Pickering 2022), nearly twice the number of species of birds or reptiles. The United States has the highest bee species richness of any country, with 3594 fully valid species (excluding territories), over half of which (1882 species) occur nowhere else (Ascher and Pickering 2022) and are of great interest for conservation. However, the sheer number of species poses an immense challenge to documenting their distributions. We must identify the gaps and spatial biases in bee biodiversity data to inform conservation planning and develop efficient strategies for minimizing biodiversity loss (Girardello et al. 2019). Results from completeness analyses can reveal the geographic areas and taxonomic groups that lack sufficient data or are prime candidates for additional analyses (Troia and McManamay 2016, Pelayo-Villamil et al. 2018, Girardello et al. 2019, Shirey et al. 2021).

Given trends in the rate of specimen digitization, it could be decades before near-complete digitization of specimen records for the majority of bee species is achieved (Cobb et al. 2019). Even if new technologies greatly increase label transcription rates (LightningBug 2022), there is no guarantee that even if all specimens were digitized, there would be enough occurrence data to estimate accurate bee species' distributions and population trends over time. Without a clearer understanding of how already-digitized occurrence records can answer basic questions about bee biogeography, we are unable to predict whether the additional specimen data locked within collections would be adequate for filling any remaining gaps about species and community trends. However, we make the case that enough information has been accumulated to assess the completeness of USA occurrence data for bees, including the 'yet-to-be-transcribed' specimen data labels. Although our assessment is only an initial step in identifying gaps and strengths in occurrence data, it can inform plans for the additional surveys and sampling initiatives required for research and conservation at ecologically meaningful scales.

We defined four study goals: 1) develop range maps for all bee species in the contiguous USA and overlap them to obtain

expected species richness patterns at any scale; 2) determine data completeness at four spatial scales for all bees and for each of the six USA bee families; 3) project the increase in completeness that would occur by including all currently undigitized specimens; and 4) identify geographic and taxonomic gaps in knowledge and provide recommendations that involve more strategic inventories and accelerated digitization. Our research provides the first comprehensive source of species ranges for bees of the United States. However, the most important contribution is our assessment of completeness for the total estimated number of specimens in USA collections and identification of areas with high richness but low sampling effort. Thus, we can identify gaps in geography and taxonomy now, even though it may be decades before already-collected specimens have their label data transcribed. The importance of conducting this type of study is underscored by the pace of global change and the growing number of projects that have documented declines in bees and other pollinators (Potts et al. 2010, Koh et al. 2016). It provides urgency for expanding on a NextGeneration set of procedures to plan strategically and execute more complete and thorough inventories (Schindel and Cook 2018).

Material and methods

We provide a more detailed methods description in the Supporting information. We acquired 2 989 647 available North American bee occurrence records from GBIF (Global Biodiversity Information Facility: 2 608 346 records) and SCAN (Symbiota Collection of Arthropod Network: 381 301 additional unique records) in February 2021 (GBIF 2021, SCAN 2021). We included all introduced species except *Apis mellifera* because its distribution in the USA is chiefly a direct result of active human management. After extensive data cleaning, including the exclusion of all undetermined morphospecies (Supporting information), our North American dataset for just the 3158 species used in our analyses was 2 157 900 records. After we removed all records for Mexico, Canada, Hawaii, Alaska, coastal islands (e.g. California Channel Islands and Florida Keys) and USA territories to restrict our data to the contiguous USA, our final data set was 1 923 814 occurrence records for 3158 bee species. All analyses described below were done using R ver. 4.0.2 (www.r-project.org), and all code can be found at https://github.com/iDigBees/USBees.

We generated observed occurrence maps and land-cover-informed range maps for 3158 described USA bee species in the contiguous USA. All 3158 bee species are listed in the Supporting information. We created occurrence maps by rasterizing the data points for a single species at a 10 × 10 km resolution and transforming them to four coarser spatial resolutions used for analyses: a) 30 × 30 km, b) 60 × 60 km, c) 110 × 110 km and d) 220 × 220 km. All individual rasters for a given resolution were then stacked to create 'observed richness' maps for the contiguous USA. These maps are presented for all bee species as a single group (Supporting information)

and for each bee family individually (Supporting information). We generated species ranges using the minimum convex polygon method by calculating the smallest polygon to encompass the full distribution of occurrence records for each bee species ('chull' function in the grDevices R package (www.r-project.org)). Since this method can over-predict species' distributions by including areas within the polygon that are not truly suitable habitat for that species, we corrected for this using 22 geospatial land cover layers (10 × 10 km resolution) from LANDFIRE (LANDFIRE 2016) to mask each species to suitable habitat (Supporting information). Cross-referencing species occurrence data with land-cover data gives an accurate classification of actual land-cover while simultaneously delineating the vegetation classes that are ecologically relevant to species distribution, thus providing a good basis for delineating species' habitat requirements. This type of approach, which distinguishes suitable from unsuitable habitat within a range, is critical for not overestimating species ranges and/or expected richness across a region and is already in common usage (Ocampo-Peñuela and Pimm 2014, Li et al. 2016). All masked species ranges for a given resolution were stacked and numerical richness values were summed across grid cells to create expected species richness maps for all USA bee species as a group (Fig. 1a, Supporting information) as well as for each family (Supporting information) at all four spatial resolutions (a–d).

Next, we ran completeness analyses for the contiguous USA by extracting numerical richness values from both observed richness and expected richness maps and calculating the ratio of observed species to expected species within a grid cell. Completeness analyses were performed at all four spatial resolutions (a–d) for all bees as a group and for each family (Supporting information). In this process, we also assessed how observational data (including records from iNaturalist and BugGuide, the top two most impactful community science websites for bees) contributed to current data completeness compared to just specimen data. Following this, we projected and incorporated ca 4.7 million additional specimen records across USA bee species into our dataset and repeated all completeness analyses to estimate the increase in completeness if we used all bee specimens known to exist in USA collections. This provides additional insights on how data adequacy (in terms of species and habitat representation) may increase if all records in museum collections were accessible after significant digitization effort. Although we expect 8 million total specimen records to reside in North American collections, 2 157 900 are already digitized and usable for the 3158 USA bee species in this study, meaning there are likely about 5 842 100 records yet undigitized. However, based on a query in SCAN, roughly 26% of those records may be composed of localities outside the contiguous USA, which brings the number down to ca 4 323 154 USA records to be digitized. We assume that the ca 376 000 USA records that were removed from analyses during data cleaning will eventually be research-ready and usable for analyses. This totals 4 699 154 additional records for USA bees that could supplement our currently usable occurrence data after being digitized or
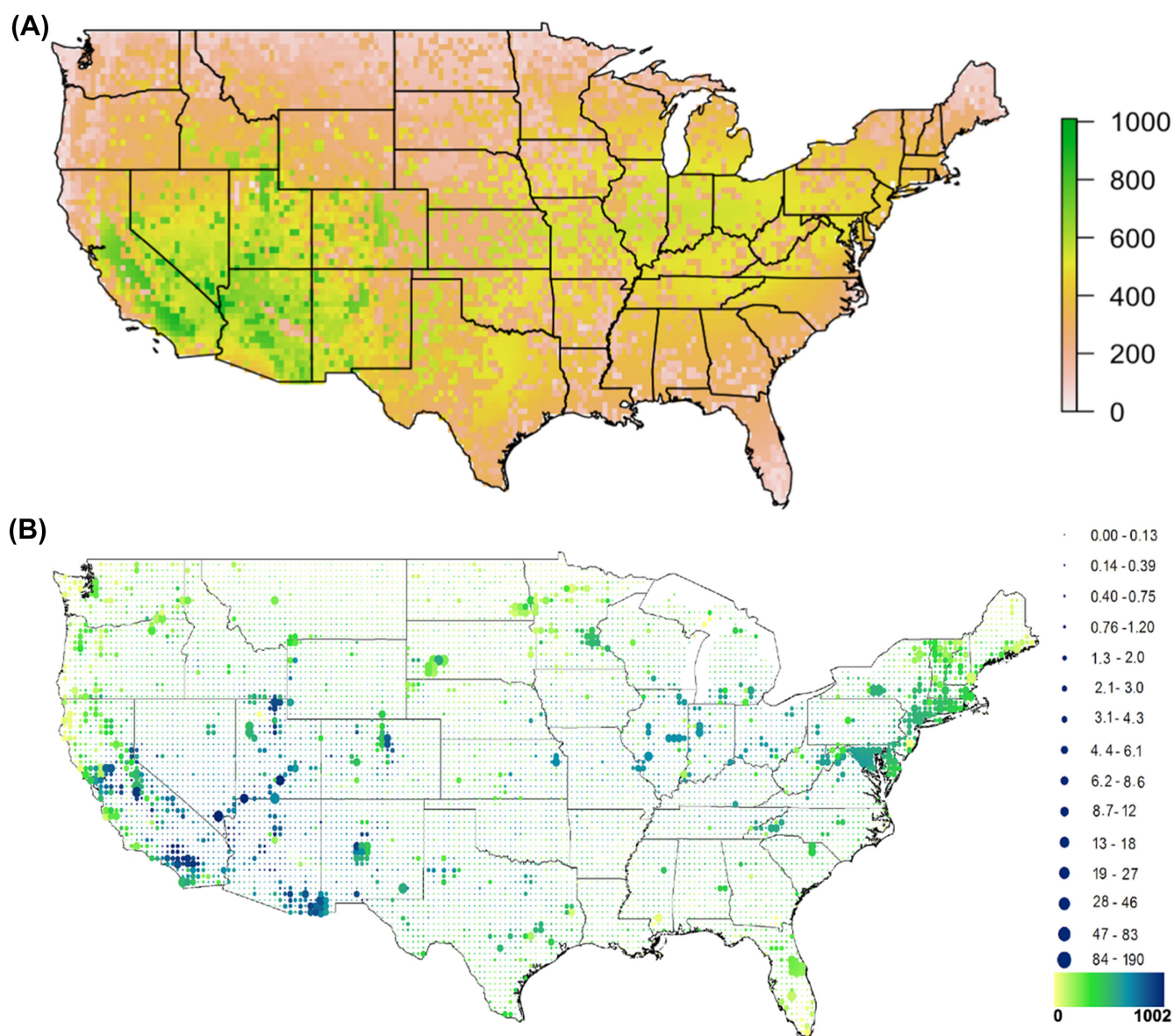
(A)

(B)

Figure 1. (A) Expected species richness for all USA bee species at a 30 × 30 km spatial resolution (900 km²). Individual species ranges were filtered by suitable habitat type using LANDFIRE land cover types and stacked to create a model of predicted richness across the contiguous USA for all 3158 bee species combined. (B) Relative sampling effort across the contiguous United States. The 'samples per species' metric represents the ratio of species to number of occurrence records per pixel. Larger circles indicate a greater number of records per species. Darker circles represent pixels with a higher number of estimated species. Pixels that are large and dark represent well-sampled areas, whereas pixels that are small and dark indicate areas with high estimated richness but have not been well sampled.

made research-ready. We rounded up and structured our code to project an even 4 700 000 additional points. New data points were proportionally allocated to each species based on the size of their masked ranges.

We used two different projection methods to generate new points. First, we distributed points within ranges and suitable LANDFIRE cover types, but spatially constrained them by existing points, through the use of a modified version of the nearest neighbor (NN) method (Clark and Evans 1954, Boyd et al. 2022). The original NN method uses the distance from an individual occurrence record to its nearest

neighbor, regardless of direction, to provide a measure of spatial relationships in populations (Clark and Evans 1954). This pattern reflects the trend that data distribution tends to be non-random, whether as a result of sampling biases or true species distributions (Boyd et al. 2022). In our adaptation, the NN values for each bee species were generated using the average nearest neighbor distance from a single record to the four closest points in each of the cardinal directions. Only unique occurrences were used when calculating nearest neighbors (unique species with unique coordinates). This method, which we refer to as the 'cardinal direction nearest neighbor'

(cdNN) method, takes into consideration the well-established trend that there are strong geographic biases in specimen-based datasets that can result in data clustering, primarily based on accessibility (Girardello et al. 2019, Hughes et al. 2021, Boyd et al. 2022). This clustering may be due to collection biases (repeat collection at field sites or accessible areas) and/or true dispersion patterns within a species' preferred range. We confirmed that the distribution of occurrences for each species is relatively clustered within its range using the occAssess R package (Boyd et al. 2021), where 99% of our species showed a significant clustering effect (i.e. values of < 1) rather than random or over-dispersed (Supporting information). We constrained the placement of projected points to fall within a circular buffer of the cdNN distance for that species. We appended all newly projected specimen records to our current North American dataset for the 3158 USA bees and regenerated observed richness maps for each species. Updated observed richness values per pixel were extracted and new completeness values were calculated for each grid cell. A histogram of the distribution of cdNN distances for all species is provided in the Supporting information.

Our second projection method used a randomized approach that projected the ca 4.7 million undigitized records anywhere within a species' range and appropriate LANDFIRE cover types without any additional spatial constraint. We again appended the newly generated points using this 'randomly projected' method to our currently digitized dataset, regenerated observed richness maps and recalculated completeness. Regardless of the projection method used, newly generated points were restricted to falling within a species' existing range polygon. We acknowledge our assumption that all ranges generated are complete, though it is unlikely that the edge of species' recorded distributions represents their absolute ranges. Importantly, however, evaluating potential range boundaries is not a goal of this paper, and we instead just 'fill in' what we already know while not extending any range boundaries.

By excluding records that correspond to < 10% of a habitat type, we could lose data from undersampled areas. However, if we did not use the masking approach, and all pixels within a minimum convex polygon were included for a species, we would likely overestimate ranges and species richness values, and thus underestimate completeness percentages. Regardless, we expect the difference in completeness values produced from the two different approaches to be minimal. To confirm the suitability of our approach, we ran separate completeness analyses where species richness values were derived from unmasked ranges and compared the difference in data completeness values at each resolution.

Lastly, we estimated the relative sampling effort across the contiguous USA for all bees and for each family. We divided the total number of current, unique records per grid cell by the total number of expected species for that same pixel to get a single value of 'samples-per-species' and compared the average 'samples-per-species' to the estimated richness in each grid cell at a 30 × 30 km resolution. We used ArcGIS (release 10.3) (ESRI 2011) to create maps that show the varying amount of samples-per-species across pixels in the contiguous USA while simultaneously showing the varying degree of species richness across those same pixels. These analyses highlight priority locations for targeted sampling with the intent of documenting species that are currently unrecorded, but expected to occur, in that area.

Results

Bee species richness across the contiguous USA

Expected bee species richness (based on overlapping range maps) and observed species richness (based on occurrence records-per-pixel) were highly variable across the United States. However, general hotspots of biodiversity are concentrated in the southwestern USA (Fig. 1a, Supporting information). When analyzed by family at a 30 × 30 km resolution, all groups except Halictidae show the same high expected diversity in the southwestern USA (Supporting information). The highest richness for any family was found in Megachilidae in the western USA (California), with 157 species currently observed and 311 species expected at a 30 × 30 km resolution (Supporting information). Patterns did not vary markedly at different resolutions for this family, though patterns around the Sierra Nevada in eastern California became much more noticeable at coarser resolutions (Supporting information). Interestingly, Halictidae showed a different richness pattern than other families, with higher expected richness primarily in the eastern USA, while the western USA showed lower and more fragmented expected richness. In general, Halictidae showed a slightly lower richness than most other families, with a maximum of 81 observed and 128 expected in the most diverse regions. Melittidae showed an even more extreme version of this pattern, with the highest richness in the far West but also patches of high richness in the East, although it should be noted that this involves a maximum of only 6 observed and 10 expected species at even the coarsest resolutions for this species-poor family. For all analyses, topographic patterns tended to become clearer as resolution became coarser (Supporting information), except for Colletidae, where trends fragmented at coarser resolutions. However, the expected richness values extracted for each pixel are likely more precise at the finer-scale resolutions due to a mismatch between landscape heterogeneity and scale at coarse resolutions.

Relative sampling effort

Sampling effort for currently digitized specimens varies considerably across the country and does not always follow patterns of species richness when species-to-sample ratios (excluding duplicates) are examined. Overall, species richness is highest in Arizona and California, but sampling effort relative to diversity was very patchy across these states. This indicates that areas already known to be species-rich may still have under-described bee faunas as they are often

proportionally undersampled, such as foothills ringing xeric regions of California. Conversely, low richness areas around Washington DC and Maryland have high numbers of samples per estimated species and therefore are less likely to have many species present that have not already been recorded (Fig. 1b). Furthermore, efforts relative to high richness are clear for certain locations such as the Southwestern Research Station of the AMNH (American Museum of Natural History) in Portal, Arizona that is used for the Bee Course (www.thebeecourse.org/ (Minckley and Ascher 2013)), and also various National Park inventories.

Importantly, family-level trends may drive underlying patterns of diversity and sampling effort (Supporting information). Andrenidae shows higher levels of sampling effort in the western USA, especially in California, Arizona and Utah, which is also an area with high bee diversity, suggesting that areas with high diversity often have disproportionately high sampling effort. However, this pattern does not hold across the entire country; there are still areas of relatively low sampling effort compared to high estimated Andrenidae richness and areas in the East where richness is lower, but sampling effort is higher. Colletidae and Megachilidae also have higher sampling effort in parts of these same western states, although bee richness is notably lower in Colletidae than Andrenidae or Megachilidae. Colletidae and Halictidae trends likely drive any high sampling efforts in eastern states; although richness for Colletidae is not notably high in the eastern USA, high samples-to-species ratios were calculated for this area. Halictidae shows the most unusual pattern, with highest richness (and most of the best sampled areas, at least for identified and digitized specimens) in the Eastern states, particularly in parts of Washington DC, Maryland, Connecticut and Delaware. Additional family trends are outlined in the Supporting information.

Current data completeness of USA bee distributions

Average data completeness for all USA bees ranges from 5.66% (± 10.23% standard deviation) at the finest resolution (30 × 30 km) to 37.57% (± 18.94%) at the largest (220 × 220 km) resolution (Fig. 2). This means that even at the coarsest spatial resolution (48 400 km), our dataset of cleaned USA bee records can only account for an average of 37.57% of all expected bee species for any grid cell in the country (Fig. 2d). When evaluated at the two middle spatial resolutions of 110 × 110 km and 60 × 60 km, mean data completeness is greatly reduced to 19.10% (± 16.19%) and 11.12% (± 12.60%), respectively. Completeness levels were similar for most families, where at the coarsest resolution, Apidae had the highest average completeness of 38.30% (± 17.03%) and Melittidae had the lowest average completeness of 29.34% (± 38.19%) (all family statistics are summarized in the Supporting information). High standard deviation in percent completeness across pixels is likely a cause of low completeness, supported by higher standard deviations for the less complete Melittidae (Supporting information).

As resolution becomes more fine-grained, completeness decreases even further, such that at a 30 × 30 km spatial resolution, even the most complete family Apidae is only 7.70% (± 11.65%) complete. Finally, to confirm our approach of using masked species ranges (convex polygons filtered by suitable land cover type) to obtain completeness percentages, we compared our results with those obtained from a separate set of analyses using unmasked range polygons. We found that between the two approaches, completeness percentages at the finest resolution (30 × 30 km) are only a maximum of 2.6% lower when using unmasked ranges, and a maximum of 0.95% lower at the coarsest resolution (220 × 220 km). These figures hold true regardless of family (Supporting information). We therefore find that our results are robust to the method of range creation.

We show family-level differences in the number of records and species contributed to the USA dataset (Fig. 3). Apidae (excluding *Apis mellifera*) contributes the most by providing 35.9% of records (690 996 records), and Melittidae contributes the least, just 0.4% of overall records (6931 records). Although Andrenidae contributes the greatest number of species (1067 species), which is 33.8% of the total species pool, it only contributes 18.3% to the total number of USA records (351 432 records), likely due to the hyper-diversity of tiny and oligolectic bees in this family that makes them challenging to collect. Completeness maps for each family are displayed in the Supporting information.

Contribution of observational records to current data completeness

There were 140 986 USA observational records contributed by iNaturalist, BugGuide and Xerces Society for 766 bee species at the time of our data pull, accounting for 24.2% of all USA species. Importantly, 74.1% of this subset of observational records correspond to the genera *Bombus* or *Xylocopa* (Apidae). Removing observational records yielded small (< 2.5%) decreases in percent completeness (Supporting information). These trends are especially driven by the removal of records in the families Apidae and Megachilidae (Supporting information).

Potential data completeness with future digitized data

Using the cdNN projection method (Supporting information), the inclusion of more digitized records (estimated to be about 4.7 million additional USA specimen records that are collected but undigitized) increased the average bee completeness to only 17.99% at the 30 × 30 km resolution (Fig. 2e). This method spatially constrains future points to maintain the clustered dispersion that most species exhibited, but with moderate capacity for future points to occur in new areas within the species' known range. Similar increases in data completeness (ranging from a 10.75% absolute increase at the 220 × 220 km resolution
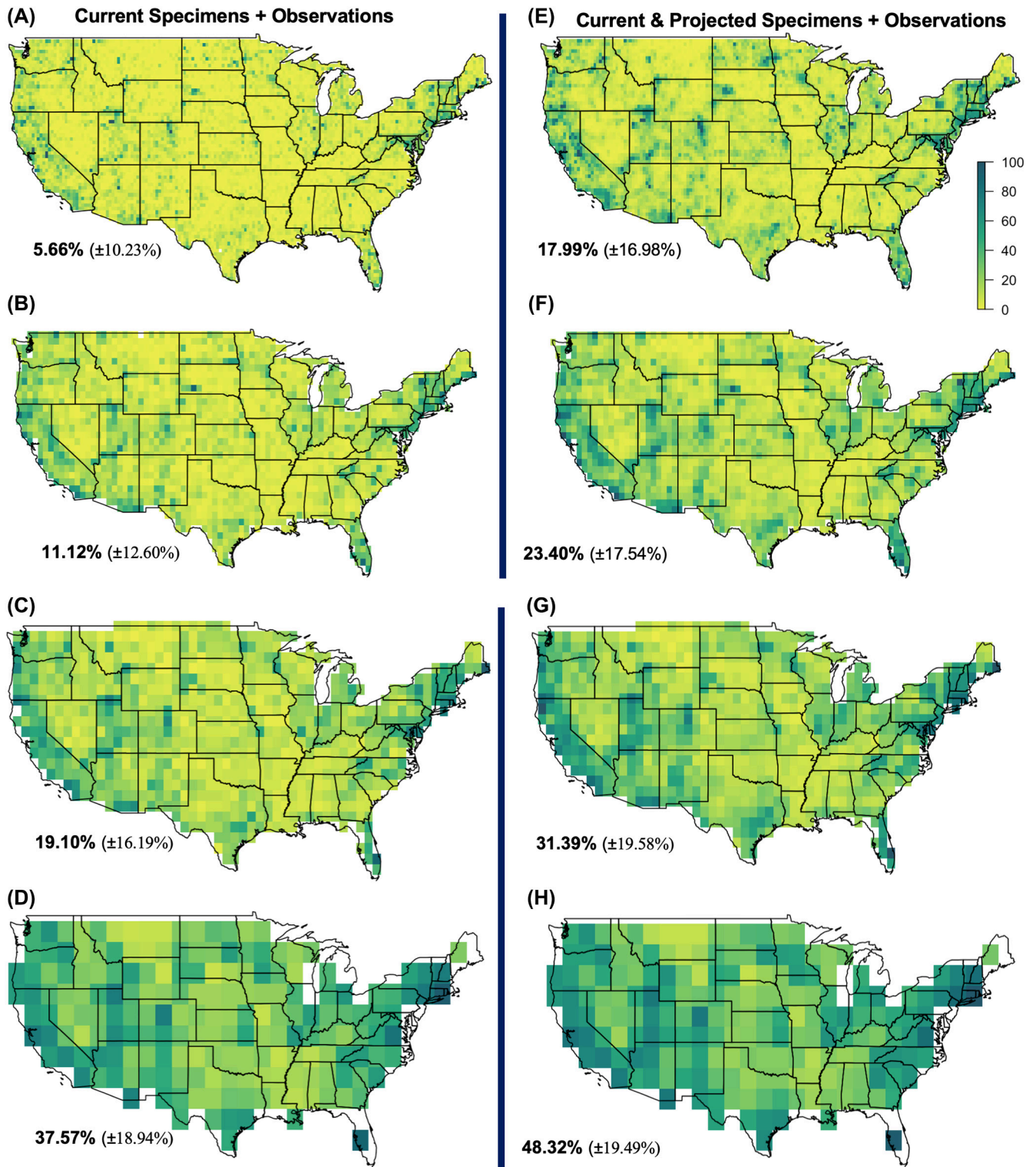
Figure 2. Percent data completeness across the contiguous USA for all 3158 bee species using current specimens and observations (A)–(D) and using a larger dataset with 4.7 million additional projected occurrences generated from the cardinal-direction-nearest-neighbor-method (E)–(H). (A) and (E) = 30 × 30 km resolution, (B) and (F) = 60 × 60 km resolution, (C) and (G) = 110 × 110 km resolution, (D) and (H) = 220 × 220 km resolution. Average percent completeness and standard deviation across all pixels is presented below each map.

to a 12.29% absolute increase at the 110 × 110 km resolution) were calculated for the other three spatial resolutions (Fig. 2f–h). Analyses at the family level indicate similar percent increases, ranging from a 7.69% absolute increase for Halictidae at a 220 × 220 km resolution to a 16.11% absolute increase for Melittidae at a 220 × 220 km resolution (Supporting information).

Results from the 'randomly projected' method, which allows points to fall anywhere within a species' current range, obtained much higher completeness values (Supporting information). Even at the finest 30 × 30 km spatial scale, overall bee data completeness increased to 86.86% for the contiguous USA when future projected records were placed randomly. At the coarsest spatial resolution, average data completeness approached 100% at 99.15% complete.

Discussion

We addressed all four goals of the study: 1) developed range maps for 3158 documented bee species occurring in the contiguous United States to create expected species richness; 2) determined low to moderate inventory completeness for bee assemblages across the USA at four spatial scales; 3) evaluated the additional contribution by community science data, and more importantly, the small increase in completeness when projecting all specimen records from USA collections; and will 4) provide recommendations for filling gaps that will lead to more complete knowledge of species assemblages to inform future inventories and monitoring for basic research and conservation efforts.

The most important implication of this research is likely the potential limitation of specimen data to understand the distribution of species and communities. This is the first study we are aware of that assesses inventory completeness for a major insect clade using existing occurrence data to project untranscribed specimen label data. The two projections from yet-to-be-transcribed specimen data were telling; the constrained projection, where points were limited by the proximity of existing bee data using cdNN values, only yielded moderate completeness increases (12–13%). This underscores the bias towards collecting around roads, municipalities, field stations or areas of geographic appeal (e.g. nature reserves), leaving large gaps in our documentation of species ranges (Meyer et al. 2015, Girardello et al. 2019, Jamieson et al. 2019, Hughes et al. 2021, Shirey et al. 2021). It is likely that the remaining specimens needing digitization will also be highly biased and this supports the notion that even if all bee specimens in USA collections (8 million) are digitized, the data will still be inadequate to provide completeness assessments at meaningful ecological scales. However, this must be tested by expeditiously digitizing all the remaining specimens. Unfortunately, it may be decades before the remaining 4.7 million bees with United States-based localities hosted in USA collections are fully digitized (Cobb et al. 2019), but we need to increase data completeness now in order to inform conservation research efforts. The second projection assumed random allocation of new points anywhere throughout a
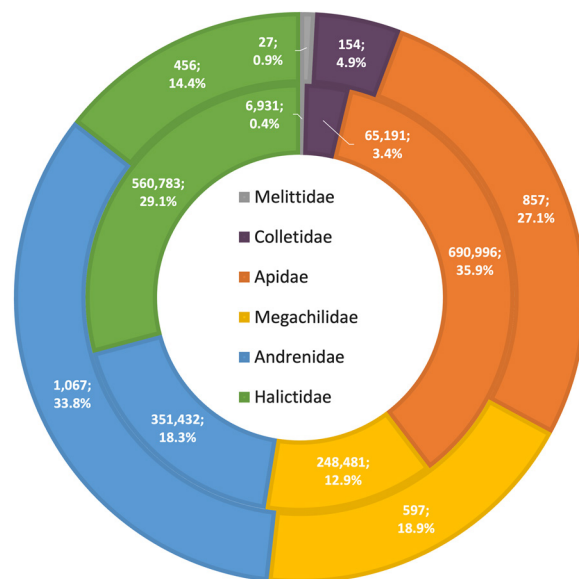

Figure 3. Inner ring = number and percentage of occurrence records contributed by each bee family for the contiguous USA data. Outer ring = number and percentage of species contributed by each bee family for species in the contiguous USA.

species' known range and generated estimates of near complete coverage (95–100%), suggesting that completeness is more limited by highly-biased sampling (Hughes et al. 2021) rather than by the number of specimens collected. This is promising in that obtaining a more complete inventory will likely only require a fraction of specimens collected to date.

The spatial biases from both specimen and observation data will require filling gaps through new inventories and backcasting where possible; continued digitization of specimens will also provide other benefits beyond completeness assessment. They will fill gaps at finer spatial resolutions than assessed here, are vital for obtaining consistent taxonomic coverage across years (Boyd et al. 2022), provide baseline historical reference (Bartomeus et al. 2019), and can serve as reference material for specimen identification, identification keys and state/regional checklists (Jamieson et al. 2019). Regardless, we need to develop inventory strategies to fill known taxonomic and geographic gaps. By extending a NextGen philosophy that embraces emerging technology and strategic planning, prioritizing sampling locations and obtaining complete collection data, including biotic associations, genetics and field images (Schindel and Cook 2018), natural history collections can play a huge role in this process by actively enabling integrated efforts that include community science (i.e. image monitoring/inventory), DNA barcoding and collaborating on open-source data projects that track progress in filling these knowledge gaps.

Spatial and taxonomic gaps

Our results uncover clear sampling biases and show that many potential hotspots of richness have low degrees of inventory completeness. For example, a recent study has shown that

the San Bernardino Valley (Arizona) exhibits the highest bee species density in a limited area of 16 km^2 compared to any other area in the world with a total of 497 bee species, constituting roughly 14% of USA species (Minckley and Ascher 2013, Minckley and Radke 2021). While our results indicate higher expected diversity in this area, digitized sampling effort is incommensurate and data completeness remains low. Other areas in the Southwest have recorded very high species richness (Michener 1979, Carril et al. 2018, Meiners et al. 2019, McCabe et al. 2020, Orr et al. 2021), though many of these areas still have a low number of records-per-estimated-species and bee data are not as complete for this region as one might expect based on our species richness maps. There are likely still additional undiscovered species and range extensions even in these already diverse regions.

Geographic and taxonomic biases are especially apparent in our expected richness map for Halictidae. Although our results predict the highest expected richness to occur in the northern Midwest, it is likely that the western USA actually has more halictid species than are reflected in our analyses. Discrepancies with the level of recent taxonomic work in certain areas could be responsible for this pattern. For example, the eastern species of *Lasioglossum*, the most speciose genus of Halictidae, have received relatively more taxonomic attention (Grundel et al. 2011, Gibbs et al. 2013, 2017). Conversely, many *Lasioglossum* specimen records collected in western state checklists were excluded from our analyses due to only being identified to morphospecies (Carril et al. 2018, Jamieson et al. 2019, McCabe et al. 2020). This could potentially explain the lower Halictidae richness in the western USA. However, recent and extensive revisions on the western *Lasioglossum* (*Dialictus*) (Gardner and Gibbs 2020) will ideally allow for these unidentified specimens to receive species identifications and get updated on data portals.

Our results corroborate other recent assessments of biodiversity data completeness: data are heavily biased (Girardello et al. 2019, Orr et al. 2021, Kass et al. 2022). A recent study of North American butterfly occurrence data (Shirey et al. 2021) identified geographic gaps across the far north, midwestern United States and northern Mexico, along with noticeable under-sampling in desert, tropical and boreal-arctic regions. For bees, a 2014 assessment of worldwide distributional data completeness for 5836 species (ca 30% of global species) indicates that global bee survey effort is unevenly distributed, with western North America and central and northern Europe having the most records (Lobo et al. 2018). However, authors of some of these assessment papers acknowledge that the country-level completeness analysis they performed does not address smaller-scale assessments on inventory quality. Essentially, areas of low inventory completeness might be masked by areas of high inventory completeness, especially in large countries such as the United States. Our study is unique as it uses 30 × 30 km as the finest-grain spatial resolution, enabling a better view of the pixels with very incomplete data that may have otherwise been masked at coarser spatial scales, especially in more environmentally-heterogeneous areas (Lobo et al.

2018). In summary, while we show that some areas within the southwestern (California, Arizona, Utah), northeastern (Maryland, Delaware, New Jersey, Vermont) and midwestern (Michigan, Illinois, North Dakota, South Dakota) United States do have small patches of higher samples-per-species, most of the country has much lower sampling effort, including patches within the states mentioned. This is especially true for Arizona and southern Nevada, despite the high potential diversity. We also show noticeable under-sampling in the south-central and mid-south regions of the USA, with less than 0.76–1.20 samples-per-species. The initiation of future inventory and monitoring programs and/or increased digitization efforts for previously-collected specimens is especially critical for the southern and central United States.

Advancing bee systematics and identification

Our completeness results are based on current knowledge. As such, our inferences are limited in part by what remains to be done for North American bee systematics and the ability of researchers to identify species properly. Many species-rich USA bee genera, representing well over 600 species, have not been included in published taxonomic revisions for decades. For example, *Nomada* and *Sphecodes* have never had comprehensive USA revisions. The situation is also challenging for Melittidae, where the largest genus, *Hesperapis*, has an inaccessible manuscript revision (Stage 1966). Similarly, *Dufourea* (Halictidae) and *Stelis* (Megachilidae) have near-finished revisions. Even for taxa that have been recently revised, identifications are challenging. For recently revised genera, such as *Epeolus* (Onuferko 2018), there has been little time for identifications of museum specimens to be made, but myriad formerly obscure or new taxa have been recognized by the author (Onuferko) and others on iNaturalist and BugGuide. Undescribed species and/or specimens from difficult-to-identify groups also pose a challenge; for example, undescribed morphospecies still comprise 16–30% of species in the western USA for projects that have had material examined by taxonomic experts (Carril et al. 2018, Delphia et al. 2019, Meiners et al. 2019, McCabe et al. 2020). To make additional progress, it is critical that taxonomic efforts are better funded, and, going forward, better integrated with emergent technologies (e.g. DNA, AI identification) that will empower ecologists and community science efforts to contribute and use integrative approaches to clarify the status of species.

Strategically filling gaps

Integrating sampling efforts
Our sampling effort analyses show that specific collections or institutions can make huge contributions to overall knowledge; some of the best-represented areas are due to extensive park survey efforts by BBSL (Bee Biology and Systematics Laboratory) or massive digitization events, including the digitization of over 369k bee specimen records through the AMNH (Ascher 2016). Additionally, geographic areas surrounding certain universities and specific researchers'

field sites are frequently the best represented (e.g. Laurel, Maryland). To even out geographic biases country-wide, we recommend supporting existing and new community science partnerships across these institutions or universities. Community science data can already contribute to overall biodiversity data completeness on local and regional scales; for example, observational data are increasing butterfly data completeness across the continent (Shirey et al. 2021). Bee records on iNaturalist are fast approaching one million for the USA alone and already cover more than one thousand species (iNaturalist.com 2022). This number is rapidly growing as established taxonomic experts engage with the site and as emerging experts increase in skill. While our results indicate that observational data only increased bee data completeness by 1–2%, this is likely an underestimate since only ca 34% of existing iNaturalist records were included in our data pull, whereas the remaining records were not identified to species and/or not considered Research Grade. Additionally, taxa such as *Bombus* and *Xylocopa* are disproportionately represented in our observational data. We are confident that the number of iNaturalist records that are identified to species will increase as image detection, combined with integration of date and location, is perfected. More importantly, it will be critical to promote concerted programs that move beyond urban centers and sample more USA locations. Thus, with continued investment by local and regional experts in identifying and curating observational records, both existing and future community scientist observational data can hugely contribute to our understanding of USA bee distributions, especially in areas where formal collections might be difficult (e.g. urban areas). The implementation of 'NextGen' collecting and curating practices (Schindel and Cook 2018), including organized monitoring programs for undersampled locations and taxa across the country, inclusive of governmental efforts and open data publication, can reduce biases in bee data. Furthermore, we can expand upon state and regional bee checklists; while there are a few USA states that have published region-specific bee species checklists (Scott et al. 2011, Carril et al. 2018, Stephenson et al. 2018, Delphia et al. 2019, Kilpatrick et al. 2020, 2021, McCabe et al. 2020, Wright 2021, Veit et al. 2022), several more states with ongoing monitoring of species occurrences (Dibble et al. 2017, Kilpatrick et al. 2020, 2021) and a few federal agency lists (Meiners et al. 2019), these are generally lacking. We encourage more of these state-level checklists, especially for the critically undersampled states comprising the central and southern USA, and the establishment of long-term local monitoring programs to help fill in geographic gaps in states where we expect high diversity yet still show noticeable under-sampling (e.g. parts of California, Arizona, Utah and New Mexico). Additionally, undescribed morphospecies present major roadblocks to capturing complete bee data, even in existing state checklists. If we can complement monitoring programs (Woodard et al. 2020) with technological advances such as DNA barcoding and other molecular tools, we can potentially reduce the number of morphospecies listed in state checklists and enhance

accuracy in species-level identification for collected bees (Jamieson et al. 2019).

Lessons from the bees: navigating the ways to meet the challenges of incomplete data on other invertebrate taxa and regions

Here we showed that, despite intense effort and taking care to integrate synonyms and remove inaccurate records (which requires expert knowledge and is hard to automate), substantial knowledge gaps remain for bees. These gaps include species that might have been locally extirpated and have not been re-recorded in many decades and/or a misalignment between sampling effort and richness in many regions, likely due to many bee species' preference of desert or mountainous habitats where human population is often low (Orr et al. 2021). For bees, undigitized data are not projected to enhance completeness percentages dramatically, but this may not be true for other taxa and/or regions for which sampling and digitization are rarer, though the rate of increase for completeness is likely to plateau for any taxon. Thus, regardless of taxonomic group, more efforts are needed to include existing specimens that are not recorded in online repositories. However, over 17 million records already exist for approximately 140 000 USA taxa (SCAN), indicating that there are enough data across different groups for the techniques highlighted in our study to be applied to other taxa.

Conclusion

Currently, the USA has the most raw bee occurrence data of any country in the world with over 60% of global bee database samples and 17.5% of all bee species documented (Orr et al. 2021). While we could map probable richness at a high resolution, our results confirm that currently available data are incomplete across space and taxonomic groups. Community science data and yet-to-be-transcribed specimen label data are still critical but will not increase completeness to the degree needed to provide a baseline across space and species that could guide strategic conservation and management. Our analyses revealed not only diversity patterns within specific bee families but also geographic areas where current samples are likely to under-represent the richness of species estimated to be present there. Conversely, most of the well-studied hotspots are limited to a few highly studied sites and regions in the west and southwest, with more work needed over the larger region, while much of the east, except the southeast, is already well-documented. Targeted and complete bee inventories and monitoring programs on a continental scale could provide the platform for future analyses with critical ecological and conservation implications, such as correlating bee species distributions with flowering plant species or nesting requirements and managing landscapes to meet these needs. Ultimately, our results provide a foundation for future monitoring programs that can be applied across the United States and increase bee data completeness on a

regional scale, while acting as a model for other countries to enact pollinator protection programs. We are confident that we can justifiably engage all collections and state and federal inventory programs with the aforementioned opportunities for strategically filling gaps to facilitate the research and management needed to conserve species and their ecosystem services in the future.

Author contributions

Paige R. Chesshire: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (lead); Writing – review and editing (lead). **Erica E. Fischer**: Data curation (supporting); Formal analysis (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Nicolas J. Dowdy**: Formal analysis (supporting); Methodology (supporting). **Terry L. Griswold**: Conceptualization (supporting); Methodology (supporting); Writing – review and editing (supporting). **Alice C. Hughes**: Conceptualization (equal); Formal analysis (equal); Investigation (supporting); Methodology (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Michael C. Orr**: Conceptualization (equal); Investigation (supporting); Methodology (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **John S. Ascher**: Investigation (supporting); Writing – review and editing (supporting). **Laura M. Guzman**: Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Keng-Lou Hung**: Data curation (supporting); Investigation (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Neil S. Cobb**: Conceptualization (equal); Data curation (supporting); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review and editing (equal). **Lindsie M. McCabe**: Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Investigation (equal); Methodology (lead); Writing – original draft (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at https://publons.com/publon/10.1111/ecog.06584.

Data availability statement

Data are available from the Figshare Digital Repository: https://figshare.com/projects/Completeness_analyses_for_over_3000_United_States_bee_species_identifies_persistent_data_gaps/138673 (Chesshire et al. 2023).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Ascher, J. 2016. Collaborative databasing of North American bee collections within a global informatics network project. – iDig-Bio Darwin Core Archive Recordset, www.idigbio.org/portal/recordsets/8919571f-205a-4aed-b9f2-96ccd0108e4c.

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

Bartomeus, I., Stavert, J. R., Ward, D. and Aguado, O. 2019. Historical collections as a tool for assessing the global pollination crisis. – Phil. Trans. R. Soc. B 374: 20170389.

Borsch, T., Stevens, A.-D., Häffner, E., Güntsch, A., Berendsohn, W. G., Appelhans, M., Barilaro, C., Beszteri, B., Blattner, F. R., Bossdorf, O., Dalitz, H., Dressler, S., Duque-Thüs, R., Esser, H.-J., Franzke, A., Goetze, D., Grein, M., Grünert, U., Hellwig, F., Hentschel, J., Hörandl, E., Janßen, T., Jürgens, N., Kadereit, G., Karisch, T., Koch, M. A., Müller, F., Müller, J., Ober, D., Porembski, S., Poschlod, P., Printzen, C., Röser, M., Sack, P., Schlüter, P., Schmidt, M., Schnittler, M., Scholler, M., Schultz, M., Seeber, E., Simmel, J., Stiller, M., Thiv, M., Thüs, H., Tkach, N., Triebel, D., Warnke, U., Weibulat, T., Wesche, K., Yurkov, A. and Zizka, G. 2020. A complete digitization of German herbaria is possible, sensible and should be started now. – Res. Ideas Outcomes 6: e50675.

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. and Carvell, C. 2022. Inferring trends in pollinator distributions across the Neotropics from publicly available data remains challenging despite mobilisation efforts. – Divers. Distrib. 28: 1404–1415.

Boyd, R. J., Powney, G. D., Carvell, C. and Pescott, O. L. 2021. occAssess: an R package for assessing potential biases in species occurrence data. – Ecol. Evol. 11: 16177–16187.

Carril, O. M., Griswold, T., Haefner, J. and Wilson, J. S. 2018. Wild bees of Grand Staircase-Escalante National Monument: richness, abundance and spatio-temporal beta-diversity. – PeerJ 6: e5867.

Ceballos, G., Ehrlich, P. R. and Raven, P. H. 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. – Proc. Natl Acad. Sci. USA 117: 13596–13602.

Ceballos, G., García, A. and Ehrlich, P. R. 2010. The sixth extinction crisis: loss of animal populations and species. – J. Cosmol. 8: 31.

Chesshire, P. R., Fischer, E. E., Dowdy, N. J., Griswold, T. L., Hughes, A. C., Orr, M. C., Ascher, J. S., Guzman, L. M., James Hung, K.-L., Cobb, N. S. and McCabe, L. M. 2023. Data from: Completeness analysis for over 3000 United States bee species identifies persistent data gap. – Figshare Digital Repository, https://figshare.com/projects/Completeness_analyses_for_over_3000_United_States_bee_species_identifies_persistent_data_gaps/138673.

Clark, P. J. and Evans, F. C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. – Ecology 35: 445–453.

Cobb, N. S. 2022. Global index of arthropod collections. – Online database, https://bug-collections.org/, retrieved 10 April 2022.

Cobb, N. S., Gall, L. F., Zaspel, J. M., Dowdy, N. J., McCabe, L. M. and Kawahara, A. Y. 2019. Assessment of North American arthropod collections: prospects and challenges for addressing biodiversity research. – PeerJ 7: e8086.

Delphia, C. M., Griswold, T., Reese, E. G., O'Neill, K. M. and Burkle, L. A. 2019. Checklist of bees (Hymenoptera: Apoidea) from small diversified vegetable farms in south-western Montana. – Biodivers. Data J. 7: e30062.

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P. and Hurlbert, A. H. 2021. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. – BioScience 71: 1179–1188.

Dibble, A. C., Drummond, F. A., Stubbs, C., Veit, M. and Ascher, J. S. 2017. Bees of Maine, with a state species checklist. – Northeast. Nat. 24: 1–48.

ESRI 2011. ArcGIS Desktop: release 10. – Environmental Systems Research Inst.

Gardner, J. and Gibbs, J. 2020. The 'red-tailed' *Lasioglossum* (*Dialictus*) (Hymenoptera: Halictidae) of the western Nearctic. – Eur. J. Taxon. 725: 1–242.

GBIF 2021. Global Biodiversity Information Facility. – Online database, www.gbif.org, retrieved 1 February 2021.

Gibbs, J., Ascher, J. S., Rightmyer, M. G. and Isaacs, R. 2017. The bees of Michigan (Hymenoptera: Apoidea: Anthophila), with notes on distribution, taxonomy, pollination and natural history. – Zootaxa 4352: 1–160.

Gibbs, J., Packer, L., Dumesh, S. and Danforth, B. N. 2013. Revision and reclassification of *Lasioglossum* (Evylaeus), L. (*Hemihalictus*) and L. (*Sphecodogastra*) in eastern North America (Hymenoptera: Apoidea: Halictidae). – Zootaxa 3672: 1–117.

Girardello, M., Chapman, A., Dennis, R., Kaila, L., Borges, P. A. and Santangeli, A. 2019. Gaps in butterfly inventory data: a global analysis. – Biol. Conserv. 236: 289–295.

Goulson, D. 2019. The insect apocalypse, and why it matters. – Curr. Biol. 29: R967–R971.

Gray, A. 2019. The ecology of plant extinction: rates, traits and island comparisons. – Oryx 53: 424–428.

Grundel, R., Jean, R. P., Frohnapple, K. J., Gibbs, J., Glowacki, G. A. and Pavlovic, N. B. 2011. A survey of bees (Hymenoptera: Apoidea) of the Indiana Dunes and northwest Indiana, USA . – J. Kans. Entomol. Soc. 84: 105–138.

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. and Porter, J. H. 2013. Big data and the future of ecology. – Front. Ecol. Environ. 11: 156–162.

Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C. and Qiao, H. 2021. Sampling biases shape our view of the natural world. – Ecography 44: 1259–1269.

iNaturalist.com 2022. Online database, www.inaturalist.org/observations?place_id=1&subview=map&taxon_id=630955, retrieved 1 August 2022.

Jamieson, M. A., Carper, A. L., Wilson, C. J., Scott, V. L. and Gibbs, J. 2019. Geographic biases in bee research limits understanding of species distribution and response to anthropogenic disturbance. – Front. Ecol. Evol. 7: 194.

Kass, J. M., Guénard, B., Dudley, K. L., Jenkins, C. N., Azuma, F., Fisher, B. L., Parr, C. L., Gibb, H., Longino, J. T., Ward, P. S., Chao, A., Lubertazzi, D., Weiser, M., Jetz, W., Guralnick, R., Blatrix, R., Des Lauriers, J., Donoso, D. A., Georgiadis, C., Gomez, K., Hawkes, P. G., Johnson, R. A., Lattke, J. E., MacGown, J. A., Mackay, W., Robson, S., Sanders, N. J., Dunn, R. R. and Economo, E. P. 2022. The global distribution of known and undiscovered ant biodiversity. – Sci. Adv. 8: eabp9908.

Kilpatrick, S. K., Gibbs, J., Mikulas, M. M., Spichiger, S.-E., Ostiguy, N., Biddinger, D. J. and Lopez-Uribe, M. M. 2020. An updated checklist of the bees (Hymenoptera, Apoidea, Anthophila) of Pennsylvania, United States of America. – J. Hymenopt. Res. 77: 1–86.

Kilpatrick, S. K., Gibbs, J., Mikulas, M. M., Spichiger, S.-E., Ostiguy, N., Biddinger, D. J. and López-Uribe, M. M. 2021. Corrigenda: An updated checklist of the bees (Hymenoptera, Apoidea, Anthophila) of Pennsylvania, United States of America. J. Hymenopt. Res. 77: 1–86. – J. Hymenopt. Res. 81: 181–190.

Koh, I., Lonsdorf, E. V., Williams, N. M., Brittain, C., Isaacs, R., Gibbs, J. and Ricketts, T. H. 2016. Modeling the status, trends and impacts of wild bee abundance in the United States. – Proc. Natl Acad. Sci. USA 113: 140–145.

La Sorte, F. A. and Somveille, M. 2020. Survey completeness of a global citizen-science database of bird occurrence. – Ecography 43: 34–43.

LANDFIRE 2016. Online resource, www.landfire.gov.

Leakey, R. E. and Lewin, R. 1996. The sixth extinction: patterns of life and the future of humankind. – Anchor.

Li, B. V., Hughes, A. C., Jenkins, C. N., Ocampo-Penuela, N. and Pimm, S. L. 2016. Remotely sensed data informs Red List evaluations and conservation priorities in Southeast Asia. – PLoS One 11: e0160566.

LightningBug 2022. Online database, https://lightningbug.tech/, retrieved 2 February 2022.

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

McCabe, L. M., Chesshire, P. R., Smith, D. R., Wolf, A., Gibbs, J., Griswold, T. L., Wright, K. W. and Cobb, N. S. 2020. Bee species checklist of the San Francisco Peaks, Arizona. – Biodivers. Data J. 8: e49285.

Meiners, J. M., Griswold, T. L. and Carril, O. M. 2019. Decades of native bee biodiversity surveys at Pinnacles National Park highlight the importance of monitoring natural areas over time. – PLoS One 14: e0207566.

Meyer, C., Kreft, H., Guralnick, R. and Jetz, W. 2015. Global priorities for an effective information basis of biodiversity distributions. – Nat. Commun. 6: 8221.

Michener, C. D. 1979. Biogeography of the bees. – Ann. Miss. Bot. Gard. 66: 277–347.

Minckley, R. L. and Ascher, J. S. 2013. Preliminary survey of bee (Hymenoptera: Anthophila) richness in the northwestern Chi-

huahuan Desert. – In: Gottfried, G. J., Ffolliott, P. F., Gebow, B. S., Eskew, L. G. and Collins, L. C. (eds), Merging science and management in a rapidly changing world: biodiversity and management of the Madrean Archipelago III and 7th conference on research and resource management in the southwestern deserts; 1–5 May 2012; Tucson, AZ. Proceedings. RMRS-P-67. U.S. Dept of Agriculture, Forest Service, Rocky Mountain Research Station, pp. 138–143.

Minckley, R. L. and Radke, W. R. 2021. Extreme species density of bees (Apiformes, Hymenoptera) in the warm deserts of North America. – J. Hymenopt. Res. 82: 317–345.

Nava-Bolanos, A., Osorio-Olvera, L. and Soberon, J. 2022. State of the art of knowledge of pollinators biodiversity in Mexico. – Rev. Mexic. Biodivers. 93: e933948.

Ocampo-Peñuela, N. and Pimm, S. L. 2014. Setting practical conservation priorities for birds in the western Andes of Colombia. – Conserv. Biol. 28: 1260–1270.

Onuferko, T. M. 2018. A revision of the cleptoparasitic bee genus *Epeolus* Latreille for Nearctic species, north of Mexico (Hymenoptera, Apidae). – ZooKeys 755: 1–185.

Orr, M. C., Hughes, A. C., Chesters, D., Pickering, J., Zhu, C.-D. and Ascher, J. S. 2021. Global patterns and drivers of bee distribution. – Curr. Biol. 31: 451.e4–458.e4.

Pelayo-Villamil, P., Guisande, C., Manjarrés-Hernández, A., Jiménez, L. F., Granado-Lorencio, C., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L. and Lobo, J. M. 2018. Completeness of national freshwater fish species inventories around the world. – Biodivers. Conserv. 27: 3807–3817.

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. 2010. Global pollinator declines: trends, impacts and drivers. – Trends Ecol. Evol. 25: 345–353.

SCAN 2021. Symbiota collection of arthropods network. – Online Database, www.scan-bugs.org, retrieved 1 February 2021.

Schindel, D. E. and Cook, J. A. 2018. The next generation of natural history collections. – PLoS Biol. 16: e2006125.

Scott, V. L., Ascher, J. S., Griswold, T. and Nufio, C. R. 2011. The bees of Colorado (Hymenoptera: Apoidea: Anthophila). – Nat. Hist. Invent. Colorado 23: 1–100.

Shirey, V., Belitz, M. W., Barve, V. and Guralnick, R. 2021. A complete inventory of North American butterfly occurrence data: narrowing data gaps, but increasing bias. – Ecography 44: 537–547.

Soberón, J., Jiménez, R., Golubov, J. and Koleff, P. 2007. Assessing completeness of biodiversity databases at different spatial scales. – Ecography 30: 152–160.

Stage, G. I. 1966. Biology and systematics of the American species of the genus *Hesperapis*, Cockerell. – PhD thesis, Univ. of California.

Stephenson, P. L., Griswold, T. L., Arduser, M. S., Dowling, A. P. and Krementz, D. G. 2018. Checklist of bees (Hymenoptera: Apoidea) from managed emergent wetlands in the lower Mississippi Alluvial Valley of Arkansas. – Biodivers. Data J. 6: e24071.

Troia, M. J. and McManamay, R. A. 2016. Filling in the GAPS: evaluating completeness and coverage of open-access biodiversity databases in the United States. – Ecol. Evol. 6: 4654–4669.

Veit, M. F., Ascher, J. S., Milam, J., Morrison, F. R. and Goldstein, P. Z. 2022. A checklist of the bees of Massachusetts (Hymenoptera: Apoidea: Anthophila). – J. Kans. Entomol. Soc. 94: 81–157.

Woodard, S. H., Federman, S., James, R. R., Danforth, B. N., Griswold, T. L., Inouye, D., McFrederick, Q. S., Morandin, L., Paul, D. L., Sellers, E., Strange, J. P., Vaughan, M., Williams, N. M., Branstetter, M. G., Burns, C. T., Cane, J., Cariveau, A. B., Cariveau, D. P., Childers, A., Childers, C., Cox-Foster, D. L., Evans, E. C., Graham, K. K., Hackett, K., Huntzinger, K. T., Irwin, R. E., Jha, S., Lawson, S., Liang, C., López-Uribe, M. M., Melathopoulos, A., Moylett, H. M. C., Otto, C. R. V., Ponisio, L. C., Richardson, L. L., Rose, R., Singh, R. and Wehling, W. 2020. Towards a US national program for monitoring native bees. – Biol. Conserv. 252: 108821.

Wright, K. W., Kazenel, M. R., Bettinelli, J., Griswold, T. L., Whitney, K. D. and Rudgers, J. A. 2021. Bee species abundance and composition in three ecosystem types at the Sevilleta National Wildlife Refuge, New Mexico, USA. Ver. 2. – Environmental Data Initiative Data Portal, https://portal.edirepository.org/nis/mapbrowse?packageid=knb-lter-sev.321.2.