Leveraging local efforts to solve regional-scale ecological questions: using multiple sources of data and a multi-species occupancy model to explore bee-plant interactions

**Running title (< 45 characters):** local efforts to solve regional problems

Authors: Michelle J. Lee1☨, Graziella V. DiRenzo2☨, Yolanda Diao3, & Katja C. Seltmann3\*

1 Ecology, Evolution and Marine Biology, University of California Santa Barbara, USA

2 U.S. Geological Survey, Massachusetts Cooperative Fish and Wildlife Research Unit, University of Massachusetts Amherst, USA

3 Cheadle Center for Biodiversity and Ecological Restoration, University of California Santa Barbara, USA

☨ Both authors contributed equally to leading this study

\* Corresponding author; email xxxx

**Target journal:**

Methods in Ecology & Evolution - Research article (7000-8000 words)

**Competing interest statement**

We have no competing interests.

*This draft manuscript is distributed solely for purposes of scientific peer review. Its content is deliberative and predecisional, so it must not be disclosed or released by reviewers. Because the manuscript has not yet been approved for publication by the U.S. Geological Survey (USGS), it does not represent any official USGS finding or policy.*

*Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.*

**Abstract (350 words max)**

1: Bees and their associated interactions are declining globally, imperiling many ecosystem services they provide, such as plant pollination. Unfortunately, many bee-plant interactions are understudied, producing an incomplete picture system-level losses as a result of bee declines. Today, we have the opportunity to learn more about bee-plant interactions via opportunistic data coalesced across historic museum records, current collections, and community science initiatives in online databases, such as GloBI (Global Biotic Interactions).

2: Here, we used the GloBI database, curated, local checklists of bee and flowering plant species, phenology data, and a multi-species occupancy model alongside stochastic search variable selection to explore hypotheses related to bee-plant interactions and detection processes. Accounting for some forbidden-links or unobservable interactions, we hypothesized that bee and floral traits would impact the number and detection of interactions. We hypothesized that our multi-species occupancy model would increase our understanding of whole network structure relative to the raw interaction occurrences, reflecting a more realistic depiction of the whole network.

3: We found a strong effect of bee sociality on the probability of bee-plant interaction, where solitary bees had a lower probability of bee-plant interactions than non-solitary (i.e., social) bees. We did not find an effect of bee size, flower color, or flower shape on the probability of bee-plant interaction. We found a strong effect of source citation type and floral traits on bee-plant detection probability, where the probability of detecting a bee-plant interaction was much higher for observational citations (e.g., iNaturalist) than for collections (e.g., museums). Our modeled interaction network showed a higher level of evenness, nestedness, and specialization relative to the interaction network of raw GloBI occurrences.

4: Observations of species interactions dictate our ability to both predict and protect ecosystem structure and function. Our study is the first of its kind to utilize occupancy modeling to better understand species interactions, leveraging both aggregated, open-source databases and expert checklists. Our findings, while largely counterintuitive, stress the importance of investigating the effect of detection and collection biases on our understanding of ecological processes.

**Keywords:** imperfect detection, occupancy modeling, community science, opportunistic data, presence-only, bee-plant interactions, pollinators, stochastic search variable selection

**Introduction**

Declines in the abundance and diversity of bee species have been reported globally and regionally (Potts et al., 2010; Zattara & Aizen, 2021); however, the mechanisms of bee declines and their biological interactions with other species (e.g., plants) are poorly understood or largely unknown, making the community and ecosystem-level consequences of such declines unknown (Ramos-Jiliberto et al., 2020). Currently, our ability to evaluate the decline of the 20,000 bee species worldwide and thus their interactions relies heavily on opportunistic occurrence data found in historic museum record databases, current collections and community science initiatives (Soroye et al., 2020), which are often influenced by observer and sampling biases.

Over two billion records that document the global occurrence of organisms are available for research on platforms such as GBIF, iDigBio, iNaturalist, eBird, Dryad, GloBI, etc. These records span across all taxonomic kingdoms, and the majority of these records are obtained from singleton human observation in nature, community science initiatives, atlas data, and the digitization of historic museum records. The availability of this type of data is only increasing with the expansive use of camera traps and environmental DNA, adding new evidence of species occurrences of a species at a particular place and time in these databases worldwide (Heberling et al., 2021). Typically, the data entered into these online databases do not conform to any particular study or design, making them a type of “opportunistic” data. When using opportunistic data, there are several types of biases to be aware of, which include but not limited to: (1) taxonomic bias (e.g., particular species may be sampled more frequently than others because more is known about them or inference is desired on that species), (2) detection bias (e.g., species detectability changes over time and space as a result of observers or number of surveys), (3) spatial bias (e.g., particular locations are more easily visited), and (4) temporal bias (e.g., certain periods of time [either time of day or seasons] may be easier for samples to be collected). Although there are these caveats, several studies have successfully used opportunistic data to examine species distributions across space and time (e.g., Kéry et al., 2010), including population declines (e.g., Guzman et al., 2021) as well as range shifts and contractions (e.g., Tingley & Beissinger, 2009). Although these studies have served to better understand the impacts of climatic changes and species loss worldwide, we still lack vital ecological understanding of species interactions and the utility of these databases to explore those lines of inquiry.

Global Biotic Interactions (GloBI) is an online integrated information system for indexing and sharing structured species-interaction data (Poelen et al., 2014). GloBI is populated with indexed datasets that include species interactions. A large amount of the data from GloBI comes from interaction records obtained from museum collections and community-science projects with singleton observations via online occurrence record databases (e.g., iNaturalist). GloBI is unique in that it also includes data from research studies, such as meta-analyses and ecological studies, worldwide. The sources of the data vary in terms of objectives and study design, and the GloBI database indexes these records in a uniform manner. Like other online databases, GloBI largely consists of presence-only interaction data, meaning that some type of species interaction must have occurred for it to be documented and inputted into the database. GloBI includes the source citations of where the interaction was published with each interaction record, and geographic information if it is provided by the source. As mentioned before, GloBI only offers presence-only interaction data, leaving it up to the user/practitioner to develop their own methods for inferring the absence or non-detection of species interactions. One method to infer species non-detections is by relying on species checklists and phenology data.

Species checklists are authoritative lists of species that occur in a given geographical area, and phenology data provides insights into when a species is active at a site. Checklists are often generated by taxonomists and collectors that specialize in a particular taxonomic group, providing a forum for expert opinion on the species present at a particular site. They provide an important standard for decision-making in biodiversity conservation and land management (Johnson et al., 2020; Reyserhove et al., 2020), and checklists are usable to retroactively assign species non-detections by comparing what was observed to a list of species that are expected to occur at a site (Kéry et al., 2010). Alongside checklists, phenology data provide valuable insights into whether or not two species are active during the same period of time. By combining the use of species checklists and phenology data, we can create lists of all possible pairwise species interactions (such as which bees could possibly interact with which plants) during different periods of time at a site based on their active times; and we can infer species non-detections.

A powerful tool to analyze species detection/non-detection data are occupancy models, which separate the ecological and sampling processes and provide valuable insights to imperfect data such as opportunistic data. Indeed, occupancy models require standardized data collection and must meet specific assumptions that, if violated, may result in biased parameter estimates and ecological inference. However, researchers have come up with ways of using occupancy models for non-standardized data (such as community science initiatives) typically used for opportunistic collections (e.g., Robinson et al., 2020; Shirey et al., 2022; Sullivan et al., 2017). For example, to use historical data, authors directly re-survey sites that were surveyed in historical periods or constrained analyses to locations where two or more sampling events occurred within a calendar year (e.g., Guzman et al., 2021; Soroye et al., 2020). This replication ensures that occupancy and detection can be estimated.

Here, we explored hypotheses related to bee and flower characteristics that may drive ecological processes (i.e., what characteristics relate to a species degree of generalization?) and the detection processes (i.e., what characteristics relate to species detectability?) using a hierarchical multi-species occupancy model, stochastic search variable selection, and a combination of species checklists, trait list, interpolated(?) phenology data, and the GloBI database. Specifically, we hypothesized that the probability of a bee-plant interaction depends on bee size with larger bees having a greater probability due to greater flight strength and generalization. We also hypothesized that bee sociality would impact the probability of a bee-plant interaction because social bees tend to be more floral generalists. Sociality has been documented to coincide with greater resource use (Kaluza et al., 2017) and greater diet flexibility (Wood et al., 2018). Body size has been shown to predict pollen load diversity (Cullen et al., 2021) and larger foraging distances than smaller bees (Greenleaf et al., 2007). Lastly, we hypothesized that flower color and flower shape would also play a role in bee-plant interaction probability because yellow flowers and bowl shapes or open flowers are also more generalized than non-yellow or not bowl-shaped flowers. Biologically, these patterns will have resulted from coevolutionary adaptations between bees and plants. Pollinator traits could confer some sort of preference or pattern in floral choice.

In addition to the evolutionary forces driving patterns in bee-plant interactions, detection by both professional and volunteer researchers can be biased based on the traits of the species of interest – impacting data within other studies. First, we expect that time of year – based on seasonality and timing of collections – will affect the probability of detecting a bee-plant interaction. Next, we hypothesized that the probability of detecting a bee-plant interaction is influenced by several factors including source citation type (i.e., collection vs. observation). Several recent studies have sought to understand the quality of volunteer and community science-based data collection, and most have found that community science data can be high quality, but with proper training and standardized methodology (Kosmala et al., 2016; Kremen et al., 2011). While professional collectors have more training and experience, these efforts may still be influenced by biases that are based in vision or interest. For instance, size and coloration of both bees and plants may play an important role in observation bias. Adamo et al. (2022) found that conservation funds and efforts were targeted toward specific plant colors like blue or yellow. Flower shape may also play an important role in observer bias, as open, umbel-like flowers are easier to observe pollinator interactions than other kinds of restricted or tubular flowers. Indeed, natural history collections are often biased with an overrepresentation of charismatic fauna and flora (Adamo et al., 2021).

By leveraging additional datasets for plant and pollinator species on Santa Cruz Island, we may be able to better predict the frequency of interactions in the network based on phenology, species traits, and potential observer biases (Bartomeus, 2013). As a result, a network of model-predicted interactions may reveal interactions that are rare or less likely observed by collectors, while also accounting for some forbidden-links. These additional interactions may add to our understanding of network structure and patterns of specialization or generalization. We expect that interactions predicted by our model will reveal an interaction web more even, but less nested and specialized than our GloBI network. Both nestedness and specialization reflect system stability and resilience. These two patterns observed in networks can be overestimated – a result of species rarity or incomplete observations (Rivera-Hutinel et al., 2012). Thus, by leveraging these other datasets in our model, we may be able to predict a more accurate understanding of network structure.

**Methods**

*Data cleaning & formatting*

To explore the hypotheses related to bee and plant characteristics that relate to ecological (i.e., what characteristics relate to a species degree of generalization?) and detection (i.e., what characteristics relate to species detectability?) processes, we compiled plant and bee species checklists, bee phenology and trait data (i.e., body size, coloration, flight timing), plant phenology and trait data (i.e., flower color, bloom time), and species interaction data from Global Biotic Interactions (Poelen et al., 2014) before embarking on the statistical analysis.

To do this, we took the following steps to clean and prepare the data:

1. Defined the geographic scope of the analysis
2. Normalized species checklist names
3. Compiled bee and plant phenology and trait data
4. Normalized and cleaned the GloBI data

Below we describe what we did under each of the steps before moving onto the statistical analysis.

*(1) Geographic scope:* We defined a spatial range to explore the hypotheses based on the accessibility of authoritative species checklists for both bees and plants. We decided to define the area to a rectangular region around Santa Cruz Island, California because of the availability of both bee and plant checklists. Bees and plants have been extensively monitored and studied on the island for many years. Bee specialists Robbin Thorpe and John Asher provided the first checklist of the island in 2005 (Thorp, 2007), and Seltmann has collected extensively on the island since 2016 providing additions to the list. The plant checklist (Hasenstab-Lehman et al., 2022) is a currently unpublished list compiled by Matthew Guilliams and Kristen Lehman from Santa Barbara Botanic Garden. Both lists are based on museum specimens that include vouchered material found in natural history collections. As Santa Cruz Island is located about 20 miles off of mainland California, it is expected that the lists for this location are less prone to migration events (Schoenherr et al., 1999).

We leveraged the bee and plant species checklists to limit the possible available interactions, but utilized a greater geographic bound for our GloBI records in order to get a more accurate understanding of species interactions according to these records. Indeed, selecting a larger geographic area proved necessary once we restricted observations to those that were identified at the species level and found only 20 viable entries of bee-plant interactions for Santa Cruz Island. Thus, we defined the GloBI interaction boundaries from San Luis Obispo to San Diego, CA (latitude from 30 – 36; longitude from -150 to -116; Figure S1). We constrained these boundaries in order to include regions with similar bee and plant phenologies, given that some of these species distributions are widespread, even outside of California. In the northern part of our defined range, San Luis Obispo, CA experiences annual high and low temperatures of 47F to 71F, respectively, with an average annual precipitation of 19 inches (<https://www.usclimatedata.com/climate/san-luis-obispo/california/united-states/usca1502>); and in the southernmost part of our defined range, San Diego, CA experiences high and low temperatures of 58F to 70F, respectively, with an average annual precipitation of 10 inches (<https://www.usclimatedata.com/climate/san-diego/california/united-states/usca0982>).

*(2) Normalized species checklists:* The bee checklist for Santa Cruz Island in June 2022 had a total of 141 species. We excluded any species in our checklists not identified to species level – i.e., genus is identified but listed as “sp.”. This removed a total of three bee species (*Calliopsis sp.*, *Sphecodes sp.*, *Stelis sp.*). We also excluded *Apis mellifera* from the bee checklist because *A. mellifera* was extirpated from Santa Cruz Island in ca. 2004 (Thorp, 2007). Our final list of bee species included 139 species.

The original plant checklist consisted of 562 species. We excluded all non-flowering plant species, and excluded flowering plant genera that had no recorded bee-plant interaction in our subsetted GloBI interaction list (including plants that had an interaction with *A. mellifera*). This brought our total plant list to 150 species.

*(3) Compiling bee and plant phenology and trait data:*

*Bees*- For all bee species on the checklist, bee size was obtained from the descriptions recorded on Discover Life or from publications found on the Biodiversity Heritage Library (Gwinn & Rinaldo, 2009) (DATE ACCESSED). Bee size was recorded as a range for both males and females. Since most of the bee size metrics were correlated, for the analyses, we used minimum female size.

Bee sociality was documented using “The Bees In Your Backyard” (Wilson & Carril, 2015) including five broad categories: eusocial, semisocial, communal, nest aggregation, solitary, and parasitic. Sociality was recorded at the genus level with a 1 representing true or 0 representing false. Genera that include several nesting behaviors were recorded as true for all behaviors found within the genus.

To compile bee trait data, we used texts from Discover Life (Ascher & Pickering, 2020) and BHL and images from XX sites. To quantify bee coloration, we created binary descriptions from head to abdomen,describing the entire body of a bee (i.e., dark metallic head, stripped or dull abdomen). We used images from XX sites in order to compile these trait characterizations.

Data for bee phenology was collected using…

*Plants*- Data for plant phenology was collected using Calflora (*Calflora*, 2022). For species that did not have phenological information on Calflora, we assigned the average bloom timeframe of other listed species in the same genus. Similarly for each species, we collected data related to floral color (i.e. yellow or not) and floral shape (i.e. disc-bowl shaped corolla or not) using Calflora and Jepson eFlora (*Jepson Flora Project (Eds.)*, 2022). Floral shape refers to the accessibility of floral resources to pollinator species and follows designation by Bosch et al., (1997).

*(4) Global Biotic Interaction data normalization:* We downloaded all unique interactions with bee species from Global Biotic Interactions (Poelen et al., 2014) on 19 Nov 2020, which consisted of 304,796 observations.

*(4.1) Filtering to bee-plant interactions:* As GloBI includes interactions for all species, we first limited the dataset to bee interactions with plant species, which could be coded as a number of different interactions depending on which species is listed as the “source” and which is the “target” of the interaction. To start the data cleaning process, we standardized the columns where bee and plant species names appeared, given that they could be in either the target or source columns. After the species names were standardized, we ran the Taxonstand package in R (Cayuela et al., 2021) on the plant column to obtain the most current accepted name per plant species. We removed any rows that did not include both a bee and plant identified to species. This brought the total number of observations down to 157,893 entries.

*(4.2) Updating species nomenclature:* Next, to make sure we obtained all GloBI entries for the bees and plants on our checklists, we used a bee synonym list from Zenodo (Seltmann & Poelen, 2021) and a list that co-author K. Seltmann put together for species that did not appear in the Zenodo list to update the Globi database with the most current accepted nomenclature for each bee species.

We converted all bee and plant species names in the GloBI dataset to the most recent and agreed upon nomenclature; and then, we filtered the GloBi data to only include bee AND plant species in our checklist, which filtered down the number of observations to 7,884 entries.

*(4.3) Filtering to geographic extent:* Next, we filtered the GloBI dataset to our pre-defined spatial extent (latitude from 30 – 36; longitude from -150 to -116), which filtered our list down to 566 entries (Figure S1).

*(4.4) Cleaning source citation information:* The original GloBI dataset downloaded in Nov 2020 had a total of 40 unique source Citations. Of the 40 unique source citations, four sources were aggregated data, 22 were of collection specimens, 13 were literature, and one was from observations (Table S1).

To better understand the detection process for different source citations listed in the GloBI database, we grouped citations by contributing institutions in the Symbiota Collection of Arthropods Network (SCAN). All SCAN entries came from seven unique source citations (with the SCAN citation resolved to its constituent institutions):

[1] "American Museum of Natural History Invertebrate Zoology Collection"

[2] "Bee Biology and Systematics Laboratory"

[3] "[http://iNaturalist.org](http://inaturalist.org) is a place where you can record what you see in nature, meet other nature lovers, and learn about the natural world."

[4] "Museum of Southwestern Biology"

[5] "R. M. Bohart Museum of Entomology"

[6] "Santa Barbara Museum of Natural History Entomology Collection"

[7] "University of California Santa Barbara Invertebrate Zoology Collection"

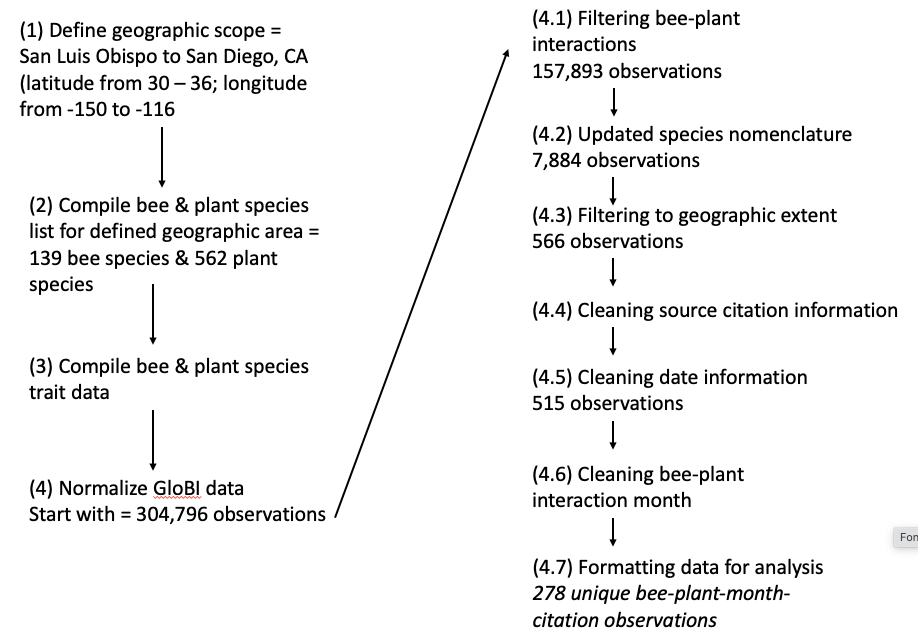
*(4.5) Cleaning date information:* Next, we formatted the event date unix epoch column in our GloBI file to year, month, and day the observation occurred. We removed all Globi observations without month information, which reduced our dataset to 515 observations.

*(4.6) Cleaning bee-plant interaction dates:* Using the compiled bee and plant phenology data, we created a list of bee-plant species interactions that were possible given the active time of bees and bloom times of plants, thus eliminating phenological “forbidden-links” (Olesen et al., 2011).

Then, we confirmed that the bee-plant interactions observed in the GloBI data file were possible against our list of possible bee-plant interactions (which was based on the phenology data). There were a few instances of a bee-plant interaction documented in the GloBI data that was not possible because it occurred outside the window of activity (45 instances). In these cases, we shifted the observation +/- 1 month to occur within the window of activity. We decided to shift the date of the observation rather than changing the phenology data because adding more months would have increased the number of possible bee-plant interactions; given the large number of already possible bee-plant interactions (>200,000).

*(4.7) Formatting data for analysis:* To format the GloBI data for the analysis, we created a 4-D array with bee species as the first dimension (139 species), plant species as the second (150 species), month as the third dimension (12 months), and source citation as the fourth dimension (7 citations). This created an array with 1,751,400 possible combinations, but based on bee-plant phenology, we found that there are only 216,385 total possible bee-plant interactions. Given that not all bee-plant interactions were possible AND that not every source citation went out into the field every month, we restricted our analysis to a total of 611,237 possible bee-plant observations within the months that each source citation was in the field.

Lastly, we populated the 4-D array with the observed bee-plant interactions using the GloBI observations. We populated the 4-D array with a value of 1 for each bee-plant interaction during the month that each source citation was in the field. We filled in the rest of the possible bee-plant interactions with a non-detection (= 0). Note that in some cases the same source citation documented the same bee-plant interaction during the same month. In total, there were 278 unique detections of bee-plant-month-citation. Also note that we collapsed all opportunistic records by source citation (e.g., iNaturalist), such that if a bee-plant interaction was ever documented, then it received a 1 (detection) in the final 4-D array, and all others were marked 0. This took care of any duplicate records in our final GloBI dataset.

**

*Model assumptions*

Given that we are using a single-level community occupancy model in an untraditional way, we explicitly outline the assumptions we are making with the data and model formulation:

1. We assume that all bee and plant species occur across our entire geographic window. Given this, an implicit assumption we are making is that bee-plant interactions are static and do not vary across space or time (i.e., no extinction or colonization events- a bee species always interacts with a plant species regardless of location and time).
2. We assume that not all bee and plant species interact with one another. We are using bee and plant phenology data to determine when each bee and plant species is active or flowering, respectively. If a bee is active and a plant is flowering, then we assume that there could be a bee-plant interaction.
3. We assume that each source citation had the opportunity to document all bee-plant interactions. We inferred non-detections for bee-plant interactions even when there is no evidence that each plant species was visited. This assumption allows us to do the following: use the different source citations as the 'repeated survey' for each bee-plant interaction.
4. We assume that all bee-plant interactions remain constant across all observations (i.e., a generalist bee remains generalist). Bees do not start interacting with new/ different plant species, and since bee-plant interactions are across time, they account for the full diet or diet changes from year to year.
5. We used source citation as a replicate ‘survey’ of each bee-plant interaction. Traditionally, replicate surveys in occupancy models are either temporal or spatial; but here, we use source citation in GloBI as independent replicate surveys, as is the case during double observer surveys.

*Model formulation*

We used a hierarchical community occupancy model to estimate the probability of bee-plant interactions and the total number of bee species interacting with each plant species (Dorazio et al., 2006). Community occupancy models are powerful tools because they preserve the identity of individual species while accounting for variable and imperfect species detection, which is not always the case for traditional methods for quantifying biodiversity (Iknayan et al., 2014; Kéry & Royle, 2016). Community occupancy models also lead to increased precision on inferences of rare and elusive species by assuming that species-specific parameters are random effects drawn from community-level distributions (Zipkin et al., 2009).

We start by defining bee species *i* interaction with plant species *j* during month *t* (Zi,j,t) as a binary variable in which Zi,j,t = 1 if bee species *i* interacts with plant species *j* during month *t*, and zero otherwise. The interaction state is assumed to be the outcome of a Bernoulli random process where:

Zi,j,t ~ Bernoulli(Ѱi,j)

Here, Ѱi,j is the probability that bee species *i* interacts with plant species *j*. We included bee size, bee sociality, flower color, and flower shape as covariates in the bee-plant interaction model using a logit-link function, where:

logit(Ѱi,j) = ui + β1 \* *Bee.sizei* + β2 \* *Bee.socialityi* + β3 \* *Flower.colorj* + β4 \* *Flower.shapej*

We standardized bee size by subtracting the mean and dividing by the standard deviation. The other three covariates (bee sociality, flower color, and flower shape) were categorical with two levels each as follows: bee sociality = solitary vs non-solitary bee, flower color = yellow or non-yellow, flower shape = bowl vs non-bowl.

We assume that each of the intercept values (ui, bee species interaction rateson the logit scale) were drawn from a community-level normal distribution with mean, μѰ, and variance, 𝜎Ѱ2, such that:

ui ~ Normal(μѰ, 𝜎Ѱ2)

The data, yi,j,t,k, consist of detection/non-detection observations. If bee species *i* is observed interacting with plant species *j* during month *t* by source citation *k*, then yi,j,t,k = 1 and a zero otherwise. True bee species interactions with plant species, Zi,j,t, are observed imperfectly, and thus, repeated sampling over a period of time when the community remains closed (i.e., no changes in bee-plant interactions) allows for the distinction between true species non-interaction and non-detection of the observation.

We define the detection model as:

yi,j,t,k ~ Bern(pi,j,t,k· Zi,j,t)

where pi,j,t,k is the probability of detecting the interaction of bee species *i* with plant species *j* during month *t* by source citation *k* given that the bee-plant interaction does occur (Zi,j,t = 1). Thus, detection is a fixed zero when a species is not present because Zi,j,t = 0 in that case. We model detection probabilities using a logit-link function where:

logit(pi,j,t,k) = vi + δ1 \* *Bee.strippinessi* + δ2 \* *Bee.sizei* + δ3 \* *Survey.montht* + δ4 \* *Survey.month2t* + δ5 \* *Citation.typek* + δ6 \* *Flower.colorj* + δ7 \* *Flower.shapej* + δ8 \* *Plant.familyj*

Again, bee size was standardized. We included a quadratic term for survey month, assuming that detection probability of bee-plant interactions peak sometime during the year and then decrease. The other five covariates (bee strippiness, citation type, flower color, flower shape, and plant family) were categorical with two levels each as follows: bee strippiness = stripped vs non- stripped, citation type: observation vs collection, flower color = yellow or non-yellow, flower shape = bowl vs non-bowl, and plant family = Asteraceae vs non-Asteraceae. We assume that each of the intercept values (vi, bee species detection probability on the logit scale) were drawn from a community-level normal distribution with mean, μp, and variance, 𝜎p2, such that:

vi ~ Normal(μp, 𝜎p2)

To estimate the total number of unique bee species interacting with each plant species (Nj), including the number of species not observed (but present) during sampling, we sum across Zi,j,t.

*Stochastic search variable selection*

To determine which covariates significantly influence the ecological and detection parts of the model above, we used stochastic search variable selection (George & McCulloch, 1993; Hooten & Hobbs, 2015). In stochastic search variable selection, we specify alternative priors for covariate *n* of the ecological model using:

β*n* | γ*n* ~ γ*n* N(0, cτ2) + (1 - γ*n*)N(0, τ2)

Both c and τ2 are tuned such that τ2 is small, providing an effective spike at zero while cτ2 is larger, creating a slab around zero. The slab then provides the prior for β*n* when the variable β*n* is in the model (i.e., when γ*n* = 1). γ*n* is interpreted as the indicator with a continuous distribution between 0 and 1. Values close to 0 suggest very little support for covariate inclusion in the model, whereas values close to 1 suggest strong support for inclusion. Given the continuous nature of the indicator, we break down values such that: 0 < γ*n* < 0.25 suggests weak support, 0.25 < γ*n* < 0.75 represents moderate support, and 0.75 < γ*n* < 1 is strong support (see indicator variable selection in Rushing et al., 2016).

We used a similar approach for the δn covariates, and we use the symbol Ωn to denote the continuous indicator (equivalent of γ*n* above) for the detection model. We also use the symbol ε2 to denote the precision parameter (equivalent to τ2 above) for the detection model.

*Model run*

We analyzed the data set using a Bayesian approach in programs R (R Core Team, 2019) and NIMBLE (package *nimble*; de Valpine et al., 2017, 2022). We ran each model for 75,000 iterations with a burn-in of 25,000 iterations and thinning by 10. We ran a total of three chains. We first ran the models with vague priors following (Lunn et al., 2012), where all parameters on the logit scale were given a normal distribution with mean zero and precision of 0.368. However, the model had trouble converging, so to aid in model convergence, we narrowed the priors based on the previous model runs. We list priors for each parameter in Table S2. We assessed convergence using the Ȓ statistic (Brooks & Gelman, 2022) and visually inspected traceplots.

*Bi-partite networks*

To better understand the difference between the resulting interactions from our model output and the raw GloBI dataset that we accessed for Santa Cruz Island, we compared interaction network metrics calculated from each bipartite matrix. To do this, we limited our modeled network to interactions that exhibited an interaction probability greater than 3%. We matched this bee species list to that observed in the GloBI lists. We created network visualizations and calculated measures of specialization (i.e. nestedness, H2’, and interaction evenness) using the package, bipartite (Dormann et al., 2009). We calculated null networks (n = 500) keeping connectance constant. We calculated the difference between observed and null values of each network metric.

**Results**

*Sample sizes*

*Checklist composition:* Our bee species included 131 solitary bee species and 6 non-solitary (or social) bees. Our bee species also included 68 stripped species and 69 non-stripped species. The median bee size was 8.5 mm (min = 2.54 mm; max = 16.5).

Our plant species included 34 species with yellow flowers, and 92 species with non-yellow flowers. Our plant species list also included 85 species with bowl shaped flowers, and 41 species with non-bowl shaped flowers. Most plant species (n = 40) were of the Asteraceae family, and a handful of other species (n = 20) were of the Fabaceae family. All other plant families had less than seven species represented in the dataset (Hydrophyllaceae[n = 6], Brassicaceae [n = 5], Euphorbiaceae [n = 5], Lamiaceae [n = 5], Onagraceae [n = 4], Aizoaceae [n = 4], etc.).

*GloBI data composition:*

*Total number of observed & estimated bee-plant interactions*

From the raw bee-plant data, there were 138 unique bee-plant interactions documented. Of those, the maximum number of observed plant interactions per bee was 19 by *Halictus tripartitus*, and the average number of plant interactions was 1.00 (SE = 0.26, range = 0 – 19; Figure 1, Table S3). The vast majority of bee species (n = 110) had 0 plant interactions documented, whereas only 27 bee species had at least one plant interaction documented in the GloBI database (Figure 1, Table S3). However, from the multi-species occupancy model, we found that the true number of unique bee-plant interactions likely ranged from 0 – 117 (mean = 6.56, SE = 1.68), where 72 bee species likely have 0 plant interactions and 65 bee species have at least 1 plant interaction (Figure 1). Based on bee-plant phenology, the average total number of possible bee-plant interactions was 116 (SE = 1.61; range = 0 – 125; Figure 1).

For a complete summary of bee-plant interaction probabilities and detection probabilities see Figures S2 & S3.

*Ecology of bee-plant interactions*

We found a strong effect of bee sociality on the probability of bee-plant interaction, where solitary bees had a lower probability of bee-plant interactions than non-solitary (i.e., social) bees (Figure 2; Table 1).

We did not find an effect of bee size, flower color, or flower shape on the probability of bee-plant interaction, where the probability these variables were included in the model was relatively low (< 0.15; Table 1).

*Detection of bee-plant interactions*

We found a strong effect of source citation type, flower color, and flower shape on bee-plant detection probability (Figure 2; Table 1). In terms of source citation type, the probability of detecting a bee-plant interaction was much higher for observational citations (e.g., iNaturalist) than for collections (e.g., museums). We also found that the probability of detecting a bee-plant interaction was higher for not yellow flowers than for yellow flowers; similarly, the probability of detecting a bee-plant interaction was higher for not bowl flowers than for bowl flowers (Figure 2).

We found a moderate effect of bee size and stripiness on bee-plant detection probability, where larger bees had a lower detection probability than smaller bees (Figure 2; Table 1) and stripped bees had a slightly lower detection probability than not striped bees.

We did not detect an effect of bee stripiness, month, or plant family on bee-plant detection probability, where the probability of these variables being included in the model was relatively low (< 0.21; Table 1).

*Bi-partite networks*

In comparing the observed GloBI network and our modeled interaction network, we found that interaction evenness of the modeled network was higher than that of the GloBI network. We found that this interaction evenness was significantly different relative to null models for the modeled network (t-value = -34.066, p-value < 0.001) and not the GloBI network. We found that the modeled network had a higher measure of nestedness relative to the GloBI network for both nestedness (t-value = -2.554, p-value = 0.005) and NODF (t-value = -9.460, p-value < 0.001). We also found that the modeled network was significantly nested relative to null models for both measures while the GloBI network was not. Finally, we found that the GloBI network had a greater level of reciprocal specialization (H2’) than the modeled network. Both the modeled (t-value = -8.859, p-value < 0.001) and GloBI networks (t-value = -6.731, p-value < 0.001) were significantly specialized relative to null.

**Discussion**

We explored several hypotheses related to bee and flower characteristics that may drive ecological processes (i.e., what characteristics relate to a species degree of generalization?) and the detection processes (i.e., what characteristics relate to species detectability?) using a hierarchical multi-species occupancy model, stochastic search variable selection, severa*l de novo* datasets (i.e., traits, phenology), and the GloBI database. We found that bee-plant interactions in our ecological model were best predicted by bee sociality, and that bee-plant detection probability was best predicted by citation type, bee size, bee coloration, and flower color and shape. We found that our modeled interaction network was relatively higher in evenness, nestedness, and specialization compared to the raw GloBI dataset. These results largely highlight detection biases that may be altering our ability to understand species interactions.

In our ecological model, we found that social bees had a higher probability of interacting with plants than non-social bees. In other words, social bees had greater generalization in their plant-partners than non-social bees. Indeed, this finding is supported by other papers that have found that social bees tend to be more generalized in their diet than non-social bees (Kaluza et al., 2017). While this pattern is well established, there remains questions about the mechanism of this relationship. For instance, non-social bees may exhibit greater floral constancy than social bees to optimize foraging strategy. Non-social bees may be targeting energy on certain resources, whereas social bees may be able to increase foraging ranges, acquiring benefits from the diversity of resources associated with a more generalized diet. This may also come about if the greater generalization of a colony of social bees exhibits floral constancy at the individual level but is – in aggregate – more generalized (Araújo et al., 2021; Bogusch et al., 2006).

This finding (i.e., social bees had a higher probability of interacting with plants than non-social bees) may have some implications for species fitness (Kaluza et al., 2018) and resilience in the context of environmental perturbations (Minckley et al., 2013) because more generalized species are able to persist in the case of plant species extinctions or losses. 86% of the bee species on the Santa Cruz Island checklist were characterized as non-social, indicating that this proportion of species could be more vulnerable in the context of ecological perturbations – such as the severe, ongoing drought in California.

Both our finding that social bees are more likely to interact with plants and our lack of evidence that bee size has an impact on these interactions may be driven by the species that make up our dataset. Indeed, the bee species, *H. tripartitus,* was the most commonly observed in our dataset and is both a social bee and a relatively smaller bee. Moreover, this bee species is found in large numbers across the island. Thus, this species may have skewed some of our results. However, some of our next most common species were larger *Bombus* species (Figure 2) and were predicted by our model to visit even more plants than that was observed. Additionally, our dataset was made up of only 14% social or eusocial bees, but that proportion does not take into account population size, which may explain why social bees have such a high probability of partaking in a plant-bee interaction.

We did not find an effect of our other covariates (i.e., bee size, flower color and shape) on bee-plant interaction probability. These unexpected results indicate that the likelihood of an interaction occurring is not necessarily influenced by bee size, flower color, or flower type. These results contrast with studies that have shown that larger bees have a proportionally larger foraging range (Greenleaf et al., 2007) and associated diet breadth (Cullen et al., 2021). Moreover, flower color and shape in other systems have been identified as important covariates to consider in plant-pollinator interactions (Bartomeus, 2013). For our study we used a comparison of yellow versus non-yellow flowers and this binary analysis could explain our lack of influence of yellow color on interaction probability. On Santa Cruz Island, we found that about 28% of plants in our dataset had yellow flowers and given this relatively minimal proportion of the dataset, we might expect that we would not have found an effect of yellow flowers on interaction probability. In contrast, 70% of plants on our plant list had bowl shaped flowers, indicating that the lack of effect of flower shape on the likelihood of an interaction was large despite the greater species richness of bowl-shaped flowers.

Santa Cruz Island, off the coast of Ventura and Santa Barbara Counties, represents a partial collection of flora and fauna found on the mainland of California. None of the bee species on our checklist are endemic species and are all found on the mainland. While there are some endemic plant species on Santa Cruz Island, there is a broad range of flora (especially at the genus level) that are present from the mainland. This is notable as bee specialization is very important at the plant genus level (Waser & Ollerton, 2006). However, there are some key differences to species makeup, including the 2004 eradication of *A. mellifera* on Santa Cruz Island. Thus, while we expect the bee-plant interactions on Santa Cruz Island to reflect similar interactions observed on the mainland due to the high similarity in genus availability we recognize that different community compositions may influence the species dynamics of our dataset.

Broadly, the results from our detection model reflect some of the behavior of collectors (both museum and community scientists). Specifically, our detection model showed that community science observers (i.e., observers using iNaturalist) were more likely to observe an interaction than the museum collector. This pattern reflects some of the obvious differences between community science and museum collections where community scientists may be more opportunistic in their data collection – recording and uploading any encountered interactions. Alternatively, museum collectors are more discerning – only recording and collecting unique interactions or recordings that address specific scientific or collection-based questions (Meineke & Daru, 2021). Thus, the differing goals of community scientists and museum collectors are likely the main drivers of these observations.

In our detection model, we also found interesting patterns in the likelihood of observing an interaction based on bee or floral characteristics. We found that smaller bees were more likely to be observed than relatively larger bees. Moreover, we found that striped bees had a similar or somewhat lower detection level than that of non-striped bees. These patterns oppose our initial predictions that larger bees and striped bees would have the highest likelihood of observation, based on the high visibility of these species. Indeed, previous studies have found that collections are usually biased towards more charismatic megafauna (Troudet et al., 2017), but it may be that collectors are targeting smaller, rarer bees rather than larger, more abundant bee species.

Finally, when we aggregated flower color to yellow and non-yellow, we did not find that yellow flowers were more likely to be visited and were not more generalized than non-yellow flowers. Similarly, when we aggregated flower shape to bowl and non-bowl flowers, we did not find that bowl-shaped flowers were more likely to be visited and were not more generalized than non-bowl flowers. These patterns may be due to the coarseness of the scale of our analysis. If we had a larger sample size, we would have parsed out flower color to all the different colors, and we would have included a larger variety of flower shapes. However, given the sample size of our dataset, we did not feel comfortable adding a large number of covariates to the model.

<BIPARTITE NETWORKS>

As we expected, model predicted interactions were much more even than in the raw GloBI network. Additional phenology and trait data revealed possible interactions unobserved by collectors for reasons that may relate to interaction or species rarity or observer biases. This was confirmed with our null networks that showed our modeled interactions were significantly even while our GloBI interactions were not. Thus, our modeled network may reveal a much richer set of interactions for each of our bee species that are not as easily observed in situ.

Contrary to our expectations, we found that the modeled network was more nested than our GloBI network and was significantly nested relative to our null models. This pattern is likely driven by the increase in predicted interactions for a subset of species including all Bombus species which may reflect the reality of the behavior of this genus which is known to be quite generalized relative to other bee genera due to body size and sociality. Thus, by leveraging information about phenology and trait matching, the model network is able to supplement our understanding of these species interactions and the structure of networks as a whole.

Finally, we observed that network reciprocal specialization (H2’) was significant for both the modeled and GloBI networks. Moreover, our GloBI network was more specialized than the modeled network. This may be a result of a few species like Bombus and Asteraceae that are most likely to interact with each other. We see that our modeled network becomes slightly less specialized, such that additional species interactions increase the overall generalization of our system.

With additional predicted networks, we see that while our networks are more nested, they may also be more generalized and even than we would predict with only observed interactions. As greater nestedness, generalization, and evenness reflect greater network stability in the context of environmental perturbations and species loss, supplemented interactions from modeled networks may reflect a more realistic and robust structure of species interactions.

While many of our results at the ecological and detection levels were unexpected, and in fact counterintuitive to what we had expected to find, we acknowledge that there are many caveats to utilizing our dataset and methodologies. For instance, when selecting the geographic region of our study to Santa Cruz Island, we included observations from the surrounding mainland to increase the overall count of interactions for our dataset. While we made this decision based on the surrounding similarity of climate and vegetation, mainland interaction observations may have other interaction partners and undergo other ecological processes outside of the context of the island system.

Throughout this course work, we discovered multiple avenues in which the study of interactions using both occupancy modeling and aggregated datasets can be improved to better predict ecological and detection approximations. Most prominently, we found that aggregated datasets can be improved with more open, discoverable, and re-usable data. The majority of our data, after filtering for necessary attributes, were largely made up of museum collections and community science observations, leaving a number of publications of recorded interactions out of our analyses. Thus, to create more re-usable data, we recommend that published records of species interactions include geolocality information and individual observations rather than collective or refined observations. A lack of these associated metadata resulted in these interactions being removed from our analysis, losing thousands of interactions from the dataset.

Adding covariates to our ecological and detection models greatly added to our understanding of both ecological and collection processes. Future studies could work toward improving these covariates, potentially leveraging valuable data regarding bee lecty or specialization to add specific probability scores to each species interaction and to remove other potential forbidden links.

<IMPORTANCE OF UTILIZING CHECKLISTS IN ECOLOGICAL STUDIES>

Our work also highlights the value of species checklists that leverage both the expertise and time of professional collectors and taxonomists.

* Taxonomic identification and observation of rare species can be difficult (Meiners et al., 2019).
* Taxonomic names require consideration
* Checklists can be found in journtals, and via GBIF
* Checklists should be cited as independent research

Observations of species interactions dictate our ability to both predict and protect ecosystem structure and function. Indeed, bee-plant interactions–specifically pollination–provide great service to whole ecosystems through the ensured persistence of plant populations that create habitat and food resources for all trophic levels. These interactions are increasingly endangered and researchers' ability to understand our own biases that shape our understanding of these interactions will prove vital in our pursuit of conservation action.

Our study is the first of its kind to utilize occupancy modeling to better understand species interactions, leveraging both aggregated, open-source databases and expert checklists. Our findings, while largely counterintuitive, stress the importance of investigating the effect of detection and collection biases on our understanding of ecological processes. Our work highlights the possibilities of using these newly populated databases and methods.

**Acknowledgments**

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

\*should include the capstone students who wrote the first version of the GloBI cleaning code?

\*institutions who assembled the checklists?

**Author contributions**

MJL contributed to project and model development, worked on data formatting and cleaning, did the bi-partite analysis, and wrote the first draft of the paper.

GVD contributed to project development, worked on data formatting and cleaning, wrote the model, analyzed the data, and contributed to editing the paper.

YD contributed to project and model development.

KS contributed to project and model development.

All co-authors edited the manuscript.

**Data Accessibility**

All data and code for analyses can be reproduced and accessed at either the github repository: XXXX or a USGS server following publication.

**References**

Adamo, M., Chialva, M., Calevo, J., Bertoni, F., Dixon, K., & Mammola, S. (2021). Plant scientists’ research attention is skewed towards colourful, conspicuous and broadly distributed flowers. *Nature Plants*, *7*(5), Article 5. https://doi.org/10.1038/s41477-021-00912-2

Adamo, M., Sousa, R., Wipf, S., Correia, R. A., Lumia, A., Mucciarelli, M., & Mammola, S. (2022). Dimension and impact of biases in funding for species and habitat conservation. *Biological Conservation*, *272*, 109636. https://doi.org/10.1016/j.biocon.2022.109636

Araújo, T. N., Pires, L. P., Meireles, D. A. L., & Augusto, S. C. (2021). Individual-resource network between Xylocopa bees and plant resources: Generalist species, specialist individuals? *Ecological Entomology*, *46*(6), 1273–1282. https://doi.org/10.1111/een.13072

Ascher, J. S., & Pickering, J. (2020). *Discover Life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila)*. https://www.discoverlife.org/mp/20q?guide=Apoidea\_species

Bartomeus, I. (2013). Understanding Linkage Rules in Plant-Pollinator Networks by Using Hierarchical Models That Incorporate Pollinator Detectability and Plant Traits. *PLoS ONE*, *8*(7), e69200. https://doi.org/10.1371/journal.pone.0069200

Bogusch, P., Kratochvíl, L., & Straka, J. (2006). Generalist cuckoo bees (Hymenoptera: Apoidea: Sphecodes) are species-specialist at the individual level. *Behavioral Ecology and Sociobiology*, *60*(3), 422–429. https://doi.org/10.1007/s00265-006-0182-4

Bosch, J., Retana, J., & Cerdá, X. (1997). Flowering phenology, floral traits and pollinator composition in a herbaceous Mediterranean plant community. *Oecologia*, *109*(4), 583–591. https://doi.org/10.1007/s004420050120

Brooks, S. P., & Gelman, A. (2022). *General Methods for Monitoring Convergence of Iterative Simulations*. 23.

*Calflora: Information on California plants for education, research and conservation*. (2022). Berkeley, California: The Calflora Database [a Non-Profit Organization]. https://www.calflora.org/

Cayuela, L., Macarro, I., Stein, A., & Oksanen, J. (2021). *Taxonstand: Taxonomic Standardization of Plant Species Names*. https://CRAN.R-project.org/package=Taxonstand

Cullen, N., Xia, J., Wei, N., Kaczorowski, R., Arceo-Gómez, G., O’Neill, E., Hayes, R., & Ashman, T.-L. (2021). Diversity and composition of pollen loads carried by pollinators are primarily driven by insect traits, not floral community characteristics. *Oecologia*, *196*(1), 131–143. https://doi.org/10.1007/s00442-021-04911-0

de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Cortes, C. W., Rodrìguez, A., Lang, D. T., & Paganin, S. (2022). *NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling* (0.13.0) [Computer software]. https://doi.org/10.5281/zenodo.1211190

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE. *Journal of Computational and Graphical Statistics*, *26*(2), 403–413. https://doi.org/10.1080/10618600.2016.1172487

Dorazio, R. M., Royle, J. A., Söderström, B., & Glimskär, A. (2006). ESTIMATING SPECIES RICHNESS AND ACCUMULATION BY MODELING SPECIES OCCURRENCE AND DETECTABILITY. *Ecology*, *87*(4), 842–854. https://doi.org/10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2

Dormann, C. F., Frueund, J., Bluethgen, N., & Gruber, B. (2009). Indices, graphs and null models: Analyzing bipartite ecological networks. *The Open Ecology Journal*, *2*, 7–24.

George, E. I., & McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, *88*(423), 10.

Greenleaf, S. S., Williams, N. M., Winfree, R., & Kremen, C. (2007). Bee foraging ranges and their relationship to body size. *Oecologia*, *153*(3), 589–596. https://doi.org/10.1007/s00442-007-0752-9

Guzman, L. M., Johnson, S. A., Mooers, A. O., & M’Gonigle, L. K. (2021). Using historical data to estimate bumble bee occurrence: Variable trends across species provide little support for community-level declines. *Biological Conservation*, *257*, 109141. https://doi.org/10.1016/j.biocon.2021.109141

Gwinn, N. E., & Rinaldo, C. (2009). The Biodiversity Heritage Library: Sharing biodiversity literature with the world. *IFLA Journal*, *35*(1), 25–34. https://doi.org/10.1177/0340035208102032

Hasenstab-Lehman, K. E., Guilliams, C. M., & Junak, S. A. (2022). *Checklist of the Vascular Flora of Santa Cruz Island, California, Version 2*.

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. (2021). Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*(6), e2018093118. https://doi.org/10.1073/pnas.2018093118

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28. https://doi.org/10.1890/14-0661.1

Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: Emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, *29*(2), 97–106. https://doi.org/10.1016/j.tree.2013.10.012

*Jepson Flora Project (eds.)*. (2022). Jepson EFlora. https://ucjeps.berkeley.edu/eflora/

Johnson, B. A., Mader, A. D., Dasgupta, R., & Kumar, P. (2020). Citizen science and invasive alien species: An analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations. *Global Ecology and Conservation*, *21*, e00812. https://doi.org/10.1016/j.gecco.2019.e00812

Kaluza, B. F., Wallace, H., Keller, A., Heard, T. A., Jeffers, B., Drescher, N., Blüthgen, N., & Leonhardt, S. D. (2017). Generalist social bees maximize diversity intake in plant species-rich and resource-abundant environments. *Ecosphere*, *8*(3), e01758. https://doi.org/10.1002/ecs2.1758

Kaluza, B. F., Wallace, H. M., Heard, T. A., Minden, V., Klein, A., & Leonhardt, S. D. (2018). Social bees are fitter in more biodiverse environments. *Scientific Reports*, *8*, 12353. https://doi.org/10.1038/s41598-018-30126-0

Kéry, M., & Royle, J. A. (2016). *Applied Hierarchical Modeling in Ecology Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. Academic Press.

Kéry, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., & Zbinden, N. (2010). Site-Occupancy Distribution Modeling to Correct Population-Trend Estimates Derived from Opportunistic Observations: Distribution Trends from Opportunistic Observations. *Conservation Biology*, *24*(5), 1388–1397. https://doi.org/10.1111/j.1523-1739.2010.01479.x

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551–560. https://doi.org/10.1002/fee.1436

Kremen, C., Ullman, K. S., & Thorp, R. W. (2011). Evaluating the Quality of Citizen-Scientist Data on Pollinator Communities. *Conservation Biology*, *25*(3), 607–617. https://doi.org/10.1111/j.1523-1739.2011.01657.x

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. CRC, Boca Raton, FL.

Meineke, E. K., & Daru, B. H. (2021). Bias assessments to expand research harnessing biological collections. *Trends in Ecology & Evolution*, *36*(12), 1071–1082. https://doi.org/10.1016/j.tree.2021.08.003

Meiners, J. M., Griswold, T. L., & 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*(1), e0207566. https://doi.org/10.1371/journal.pone.0207566

Minckley, R. L., Roulston, T. H., & Williams, N. M. (2013). Resource assurance predicts specialist and generalist bee activity in drought. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1759), 20122703. https://doi.org/10.1098/rspb.2012.2703

Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., & Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1706), 725–732. https://doi.org/10.1098/rspb.2010.1371

Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, *24*, 148–159. https://doi.org/10.1016/j.ecoinf.2014.08.005

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O., & Kunin, W. E. (2010). Global pollinator declines: Trends, impacts and drivers. *Trends in Ecology & Evolution*, *25*(6), 345–353. https://doi.org/10.1016/j.tree.2010.01.007

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramos-Jiliberto, R., Moisset de Espanés, P., & Vázquez, D. P. (2020). Pollinator declines and the stability of plant–pollinator networks. *Ecosphere*, *11*(4), e03069. https://doi.org/10.1002/ecs2.3069

Reyserhove, L., Desmet, P., Oldoni, D., Adriaens, T., Strubbe, D., Davis, A. J. S., Vanderhoeven, S., Verloove, F., & Groom, Q. (2020). A checklist recipe: Making species data open and FAIR. *Database*, *2020*, baaa084. https://doi.org/10.1093/database/baaa084

Rivera-Hutinel, A., Bustamante, R. O., Marín, V. H., & Medel, R. (2012). Effects of sampling completeness on the structure of plant–pollinator networks. *Ecology*, *93*(7), 1593–1603. https://doi.org/10.1890/11-1803.1

Robinson, O. J., Ruiz‐Gutierrez, V., Reynolds, M. D., Golet, G. H., Strimas‐Mackey, M., & Fink, D. (2020). Integrating citizen science data with expert surveys increases accuracy and spatial extent of species distribution models. *Diversity and Distributions*, *26*(8), 976–986. https://doi.org/10.1111/ddi.13068

Rushing, C. S., Ryder, T. B., & Marra, P. P. (2016). Quantifying drivers of population dynamics for a migratory bird throughout the annual cycle. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1823), 20152846. https://doi.org/10.1098/rspb.2015.2846

Schoenherr, A. A., Feldmeth, C. R., & Emerson, M. J. (1999). *Natural History of the Islands of California*. University of California Press.

Seltmann, K. C., & Poelen, J. H. (2021, May 13). *A Practical Exploration of Biotic Interaction Data Management and Information Retrieval through Terrestrial Parasite Tracker (TPT) and Global Biotic Interactions (GloBI)*. https://doi.org/10.5281/zenodo.4759060

Shirey, V., Khelifa, R., M’Gonigle, L. K., & Guzman, L. M. (2022). Occupancy‐detection models with museum specimen data: Promise and pitfalls. *Methods in Ecology and Evolution*, 2041-210X.13896. https://doi.org/10.1111/2041-210X.13896

Soroye, P., Newbold, T., & Kerr, J. (2020). Climate change contributes to widespread declines among bumble bees across continents. *Science*, *367*(6478), 685–688. https://doi.org/10.1126/science.aax8591

Sullivan, B. L., Phillips, T., Dayer, A. A., Wood, C. L., Farnsworth, A., Iliff, M. J., Davies, I. J., Wiggins, A., Fink, D., Hochachka, W. M., Rodewald, A. D., Rosenberg, K. V., Bonney, R., & Kelling, S. (2017). Using open access observational data for conservation action: A case study for birds. *Biological Conservation*, *208*, 5–14. https://doi.org/10.1016/j.biocon.2016.04.031

Thorp, R. (2007). *A List of Bees of Santa Cruz Island, CA*. https://escholarship.org/uc/item/32s13481

Tingley, M. W., & Beissinger, S. R. (2009). Detecting range shifts from historical species occurrences: New perspectives on old data. *Trends in Ecology & Evolution*, *24*(11), 625–633. https://doi.org/10.1016/j.tree.2009.05.009

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-09084-6

Waser, N. M., & Ollerton, J. (2006). *Plant-Pollinator Interactions: From Specialization to Generalization*. University of Chicago Press.

Wilson, J. S., & Carril, O. M. (2015). The Bees in Your Backyard: A Guide to North America’s Bees. In *The Bees in Your Backyard*. Princeton University Press. https://doi.org/10.1515/9781400874156

Wood, T. J., Kaplan, I., & Szendrei, Z. (2018). Wild Bee Pollen Diets Reveal Patterns of Seasonal Foraging Resources for Honey Bees. *Frontiers in Ecology and Evolution*, *6*. https://www.frontiersin.org/article/10.3389/fevo.2018.00210

Zattara, E. E., & Aizen, M. A. (2021). Worldwide occurrence records suggest a global decline in bee species richness. *One Earth*, *4*(1), 114–123. https://doi.org/10.1016/j.oneear.2020.12.005

Zipkin, E. F., DeWan, A., & Andrew Royle, J. (2009). Impacts of forest fragmentation on species richness: A hierarchical approach to community modelling. *Journal of Applied Ecology*, *46*(4), 815–822. https://doi.org/10.1111/j.1365-2664.2009.01664.x

**Figures & Tables**

Table 1. Summary of parameter estimates from multi-species occupancy model. We include parameter definitions, symbols, mean and standard deviation (SD) of the posterior distributions, and the 95% Credible Interval. All μ, β, and δ values are on the logit scale.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter definition | Symbol | Mean | SD | 95% Credible Interval | |
| Average bee-plant  detection probability | μp | -3.39 | 0.58 | -4.51 | -2.27 |
| Average bee-plant  interaction probability | μѰ | -0.51 | 0.74 | -1.99 | 0.92 |
| Standard deviation of  bee-plant detections | 𝜎p2 | 1.77 | 0.37 | 1.13 | 2.57 |
| Standard deviation of  bee-plant interactions | 𝜎Ѱ2 | 3.54 | 0.57 | 2.48 | 4.70 |
| Bee size | β1 | 1.51 | 0.63 | 0.39 | 2.85 |
| Bee sociality | β2 | -9.80 | 1.36 | -12.42 | -6.90 |
| Flower color | β3 | 1.59 | 0.43 | 0.70 | 2.42 |
| Flower shape | β4 | 1.21 | 0.43 | 0.31 | 2.02 |
| Bee size | γ*1* | 0.12 | 0.32 | 0.00 | 1.00 |
| Bee sociality | γ*2* | 1.00 | 0.04 | 1.00 | 1.00 |
| Flower color | γ*3* | 0.13 | 0.33 | 0.00 | 1.00 |
| Flower shape | γ*4* | 0.08 | 0.27 | 0.00 | 1.00 |
| Bee stripes | δ1 | -0.16 | 0.56 | -1.75 | 0.78 |
| Bee size | δ2 | -0.51 | 0.53 | -1.61 | 0.12 |
| Survey month linear | δ3 | 0.08 | 0.14 | -0.08 | 0.47 |
| Survey month quadratic | δ4 | -0.01 | 0.01 | -0.04 | 0.00 |
| Citation type | δ5 | 2.54 | 0.26 | 2.05 | 3.08 |
| Flower color | δ6 | -1.17 | 0.46 | -2.05 | -0.06 |
| Flower shape | δ7 | -1.04 | 0.48 | -1.91 | 0.00 |
| Plant family | δ8 | 0.02 | 0.12 | -0.18 | 0.35 |
| Bee stripes | Ω1 | 0.34 | 0.47 | 0.00 | 1.00 |
| Bee size | Ω2 | 0.58 | 0.49 | 0.00 | 1.00 |
| Survey month linear | Ω3 | 0.21 | 0.41 | 0.00 | 1.00 |
| Survey month quadratic | Ω4 | 0.03 | 0.17 | 0.00 | 1.00 |
| Citation type | Ω5 | 1.00 | 0.05 | 1.00 | 1.00 |
| Flower color | Ω6 | 0.94 | 0.23 | 0.00 | 1.00 |
| Flower shape | Ω7 | 0.91 | 0.29 | 0.00 | 1.00 |
| Plant family | Ω8 | 0.14 | 0.34 | 0.00 | 1.00 |

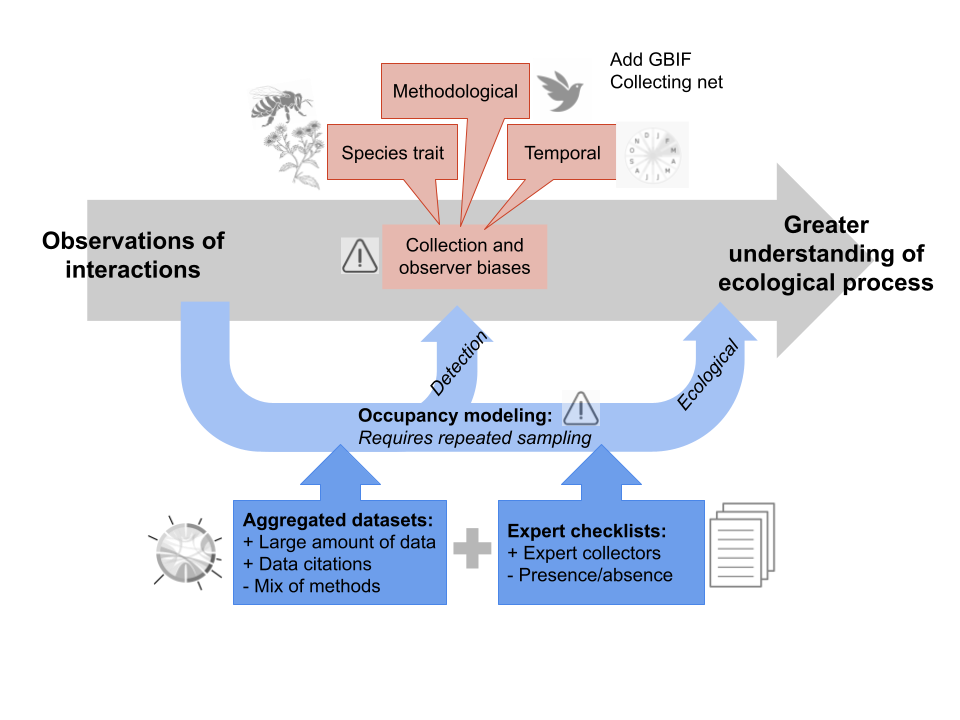


Figure 0. Conceptual model for paper. Researchers seek to better understand ecological processes through observation – in our case, through the observation of species interactions. However, our observations are fraught with collection and observer biases. In this paper we address some of these biases, including the traits of our species of interest, methodological biases, and the timing of our observations and the timing of species phenologies. In this work, we propose the use of occupancy models to estimate the likelihoods of detecting a species interaction and the ecological likelihood of an interaction by leveraging open, aggregated datasets and expert checklists.

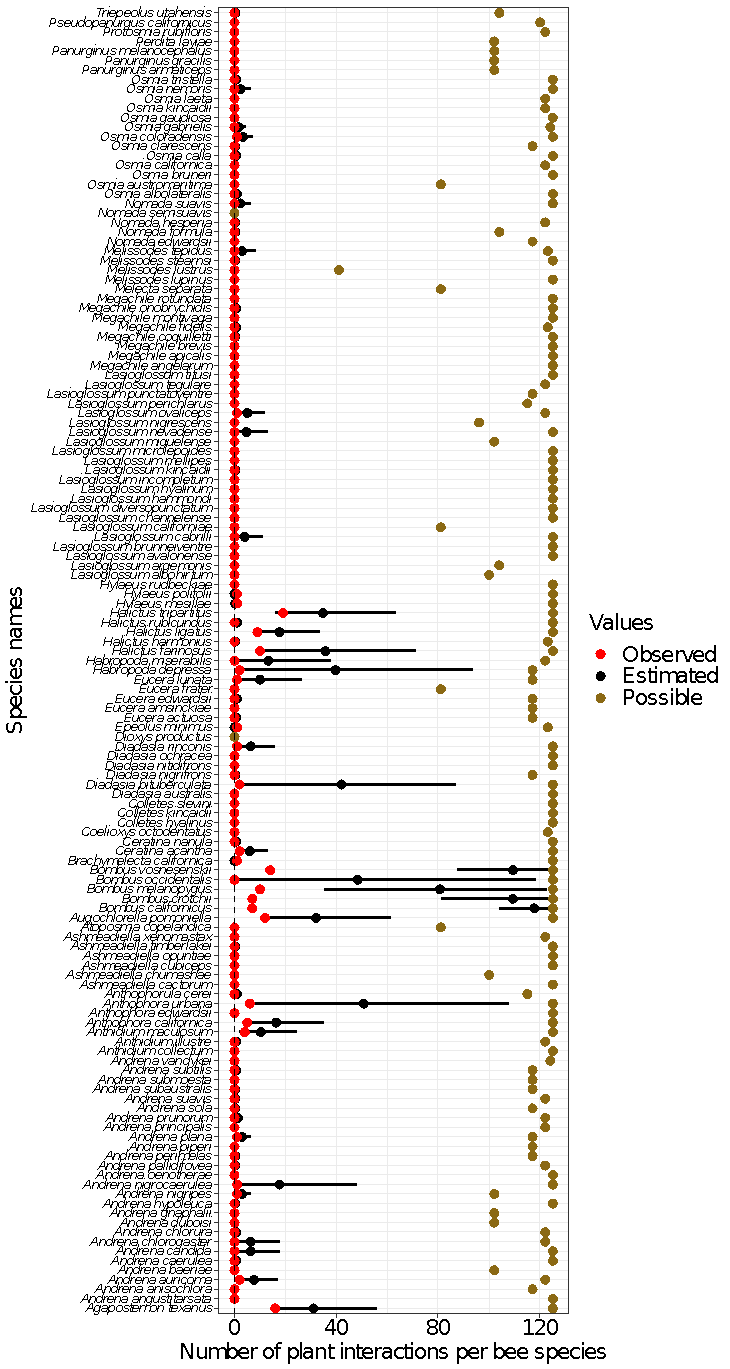


Figure 1. Number of plant interactions per bee species from the observational data (red points), the number of plant interactions per bee species estimated by the multi-species occupancy model (black points = mean; line range = 95% Credible Interval), and the number of total possible plant interactions per bee species based on phenology (yellow points). Bee species are listed in alphabetical order starting at the bottom.

Chart, diagram

Description automatically generated

Figure 2. Bee-plant interaction or detection probabilities as a function of the following covariates (A) bee sociality, (B) bee size, (C) bee stripiness, (D) citation type, (E) flower color, and (F) flower shape. The + symbols in panel (B) represent the distribution of observed bee sizes, and vertical colored dashed lines represent means to the corresponding colored distributions in each panel.

Timeline

Description automatically generated

Figure 4. Bee-plant interaction networks visualizations. (A) The GloBI network and the (B) modeled network visualizations have plants on the left-hand side ordered by order and arranged by family. Bee species on the right-hand side are ordered by family. The modeled network accounts for all interactions with a greater than 3% probability of interaction. The bee and plant species lists were matched to create the GloBI network visualization.

Supplemental Figures

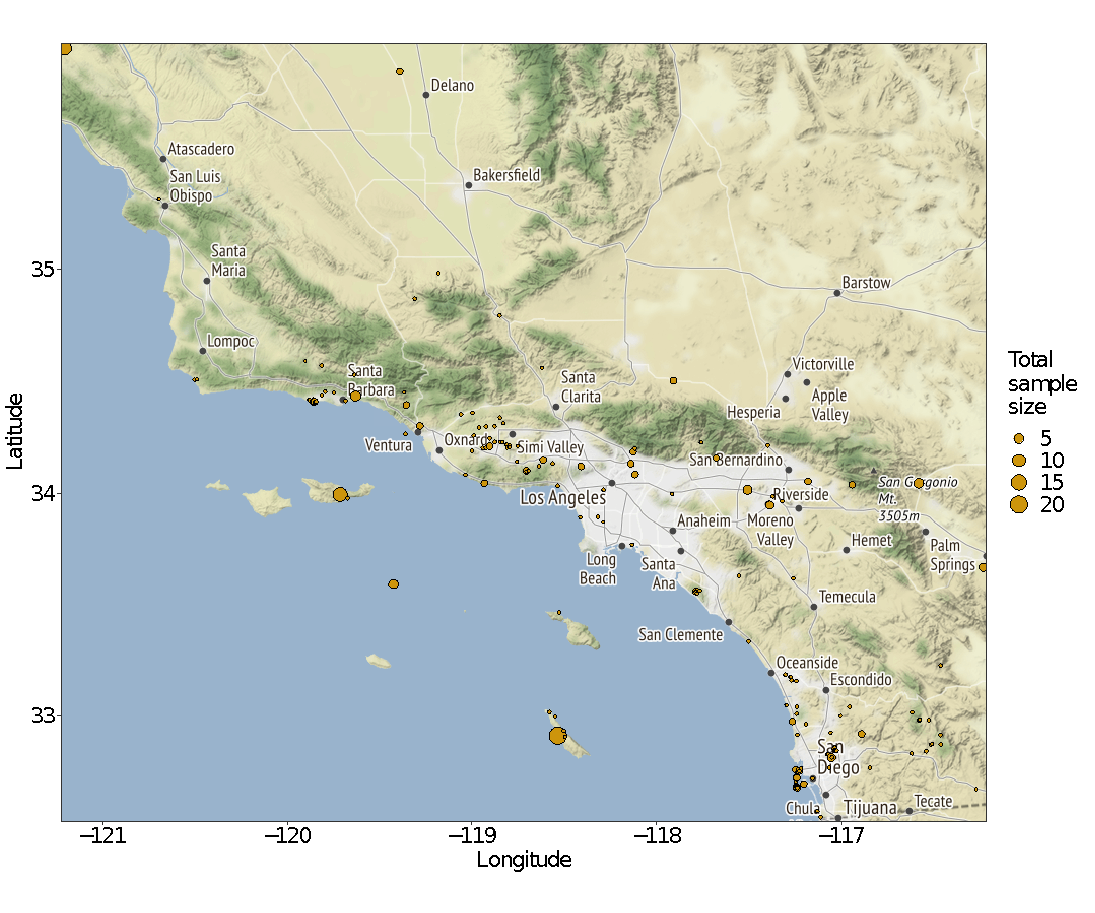


Figure S1. Geographic distribution of observations used in the analysis. Point size represents sample size at each location.

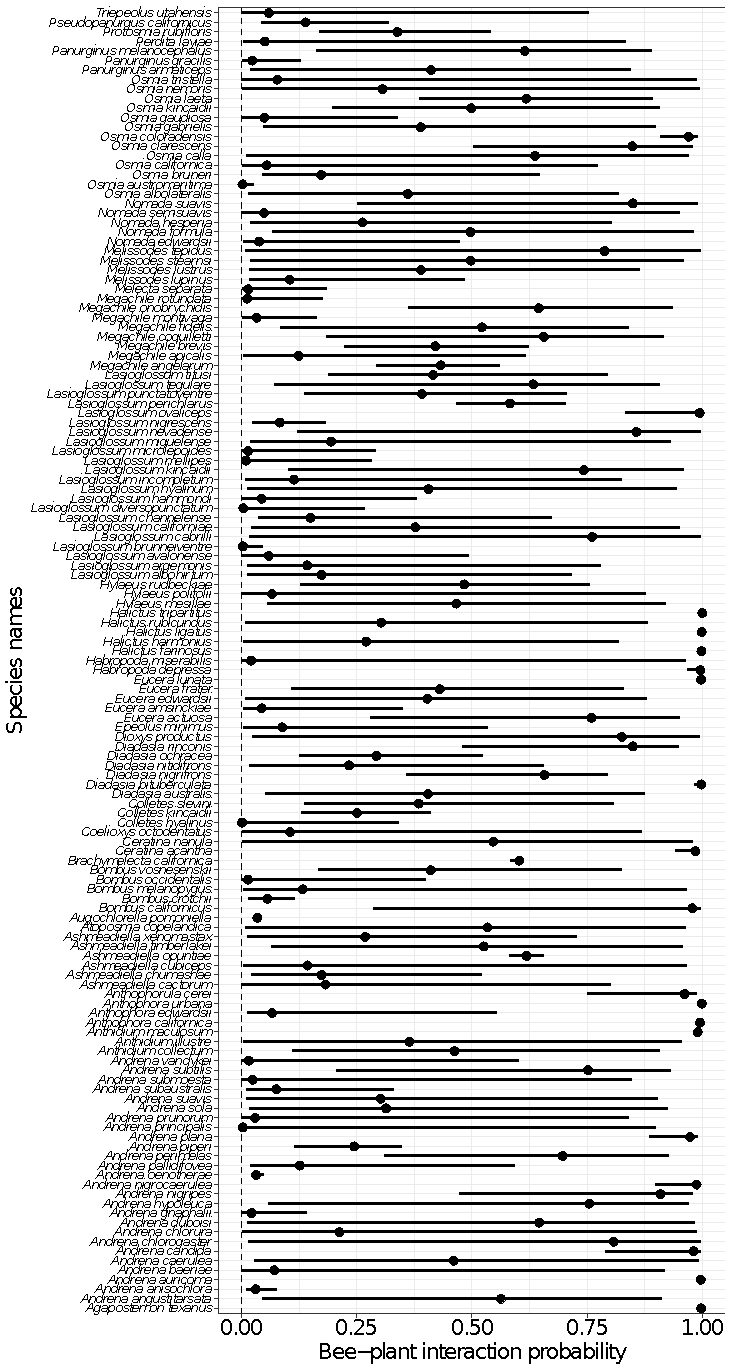


Figure S2. Summary of bee-plant interaction probabilities by bee species.

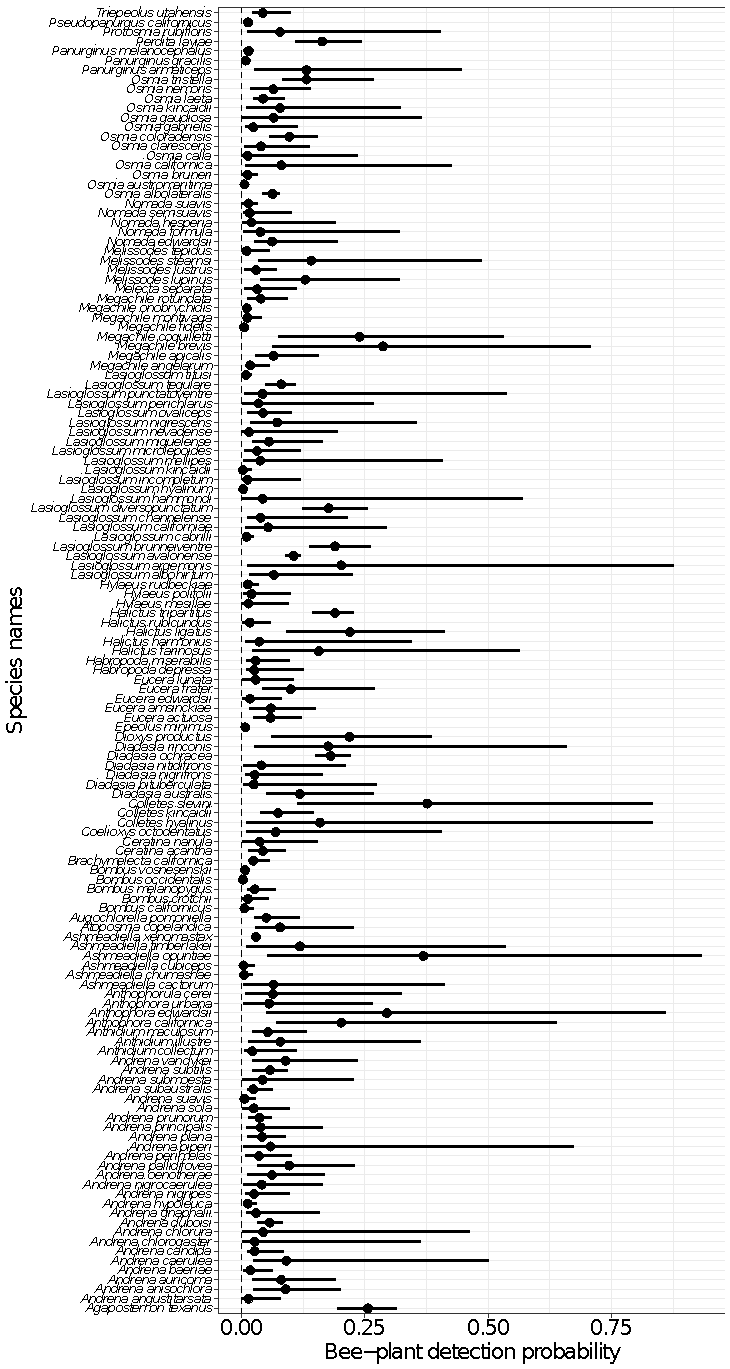


Figure S3. Summary of bee-plant detection probabilities by bee species.

Diagram

Description automatically generated

Figure S4. Heatmap visualizations of the (A) GloBI network and (B) modeled network. The modeled network accounts for all interactions with a greater than 3% probability of interaction. The bee and plant species lists were matched to create the GloBI network visualization. Bees are represented vertically, and plants are represented horizontally. The shadedness of each square represents greater frequency or likelihood of interactions.

Table S1. List of sources from the GloBI database and their reference type (aggregated data, collection specimen, literature, and observation).

|  |  |
| --- | --- |
| Source | Reference Type |
| A. Thessen. 2014. Species associations extracted from EOL text data objects via text mining. | Aggregated Data |
| Agosti, Donat. 2020. Transcription of Linne√É√Ö, C. von, 1758. Systema naturae per regna tria naturae‚Äö√Ñ√òsecundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Available at: http://dx.doi.org/10.5962/bhl.title.542 . | Literature |
| Brose, U. (2018). GlobAL daTabasE of traits and food Web Architecture (GATEWAy) version 1.0 [Data set]. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig. https://doi.org/10.25829/IDIV.283-3-756 | Aggregated Data |
| California Academy of Sciences Entomology | Collection Specimen |
| Carril OM, Griswold T, Haefner J, Wilson JS. (2018) Wild bees of Grand Staircase-Escalante National Monument: richness, abundance, and spatio-temporal beta-diversity. PeerJ 6:e5867 https://doi.org/10.7717/peerj.5867 | Literature |
| Digital Bee Collections Network, 2014 (and updates). Version: 2015-03-18. National Science Foundation grant DBI 0956388; PBI: Phytophagous Insects as a Model Group for Documenting Planetary Biodiversity (Insecta: Heteroptera: Miridae: Orthotylinae, Phylinae). Version: 08 Mar 2016. National Science Foundation grant DBI#0316495; Tri-Trophic Thematic Collection Network, 2014 (and updates). Version: 08 Mar 2016. http://tcn.amnh.org/. National Science Foundation grant(s) EF#1115081, EF#1115103, EF#1115080, EF#1115144, EF#1115191, EF#1115104, EF#1115115 | Literature |
| Eardley C, Coetzer W. 2011. Catalogue of Afrotropical Bees. http://doi.org/10.15468/u9ezbh | Literature |
| Ed Baker; Ian J. Kitching; George W. Beccaloni; Amoret Whitaker et al. (2016). Dataset: NHM Interactions Bank. Natural History Museum Data Portal (data.nhm.ac.uk). https://doi.org/10.5519/0060767 | Aggregated Data |
| el | Aggregated Data |
| Food Webs and Species Interactions in the Biodiversity of UK and Ireland (Online). 2017. Data provided by Malcolm Storey. Also available from http://bioinfo.org.uk. | Literature |
| Gandhi, K. J. K., & Herms, D. A. (2009). North American arthropods at risk due to widespread Fraxinus mortality caused by the Alien Emerald ash borer. Biological Invasions, 12(6), 1839‚Äö√Ñ√¨1846. doi:10.1007/s10530-009-9594-1. | Literature |
| Global Web Database (http://globalwebdb.com): an online collection of food webs. Accessed via https://www.globalwebdb.com/Service/DownloadArchive on 2017-10-12. | Literature |
| http://iNaturalist.org is a place where you can record what you see in nature, meet other nature lovers, and learn about the natural world. | Observation |
| https://mangal.io - the ecological interaction database. | Literature |
| https://vertnet.org | Collection Specimen |
| Illinois Natural History Survey Insect Collection | Collection Specimen |
| Jorrit H. Poelen. 2017. Species interactions associated with known species interaction datasets. | Literature |
| Katja Schulz. 2020. Collection of refuted species associations claims provided by Enclyclopedia of Life. | Literature |
| National Database Plant Pollinators. Center for Plant Conservation at San Diego Zoo Global. Accessed via https://saveplants.org/national-collection/pollinator-search/ on 2020-06-05. | Literature |
| Pensoft Darwin Core Archives available via Integrated Publication Toolkit | Literature |
| Pensoft Darwin Core Archives with associateTaxa columns | Collection Specimen |
| Purdue Entomological Research Collection | Collection Specimen |
| Quentin J. Groom. 2020. Species interactions of species on the List of invasive alien species of Union concern | Collection Specimen |
| Sarah E Miller. 12/13/2016. Species associations manually extracted from Onstad, D.W. EDWIP: Ecological Database of the World's Insect Pathogens. Champaign, Illinois: Illinois Natural History Survey, [23/11/2016]. http://insectweb.inhs.uiuc.edu/Pathogens/EDWIP. | Collection Specimen |
| Sarah E Miller. 5/30/2016. Interations from various papers. | Collection Specimen |
| Sarah E Miller. 6/19/2015. Species associations manually extracted from datasets https://www.nceas.ucsb.edu/interactionweb/resources.html. | Collection Specimen |
| Sarah E Miller. 6/22/2015. Species associations manually extracted from datasets https://www.nceas.ucsb.edu/interactionweb/resources.html. | Collection Specimen |
| Sarah E Miller. 6/25/2015. Species associations manually extracted from Robertson, C. 1929. Flowers and insects: lists of visitors to four hundred and fifty-three flowers. Carlinville, IL, USA, C. Robertson. | Collection Specimen |
| Sarah E Miller. 7/7/2016. Text gathered from Wirta, H.K., Vesterinen, E.J., Hamb‚àö¬ßck, P.A., Weingartner, E., Rasmussen, C., Reneerkens, J., Schmidt, N.M., Gilg, O. and Roslin, T., 2015. Exposing the structure of an Arctic food web. Ecology and evolution, 5(17), pp.3842-3856. | Collection Specimen |
| Sarah E Miller. 9/15/2016. Species associations extracted from http://parasiticplants.siu.edu/index.html. | Collection Specimen |
| Sarah E. Miller. 07/06/2017. Information extracted from dataset https://www.idigbio.org/portal/recordsets/db4bb0df-8539-4617-ab5f-eb118aa3126b. | Collection Specimen |
| Seltmann, Katja C. 2020. Biotic species interactions about bees (Anthophila) manually extracted from literature. | Collection Specimen |
| Semantic Prototypes in Research Ecoinformatics (SPIRE). Data provided by Joel Sachs. See also http://ebiquity.umbc.edu/get/a/publication/297.pdf . | Collection Specimen |
| Symbiota Collections of Arthropods Network (SCAN) | Collection Specimen |
| The Albert J. Cook Arthropod Research Collection | Collection Specimen |
| University of California Santa Barbara Herbarium | Collection Specimen |
| University of California Santa Barbara Invertebrate Zoology Collection | Collection Specimen |
| University of Hawaii Insect Museum | Collection Specimen |
| University of New Hampshire Collection of Insects and other Arthropods | Collection Specimen |
| Web of Life. http://www.web-of-life.es . | Literature |

Table S2. List of parameters and the priors used in the NIMBLE model fitting process.

​​

|  |  |
| --- | --- |
| Parameter | Prior |
| μѰ | dnorm(-0.25, 2.0) |
| 𝜎Ѱ2 | T(dnorm(3.75, sd = sqrt(1/2)), 0, 10) |
| μp | dnorm(-2, 2.0) |
| 𝜎p2 | T(dnorm(2.0, sd = sqrt(1/2)), 0, 10) |
| (1/τ2)1/2 | T(dnorm(1.0, sd = sqrt(1/2)), 0, 10) |
| (1/ε2)1/2 | T(dnorm(0.10, sd = sqrt(1/2)), 0, 10) |

Table S3. Summary of the number of observed vs model estimated plant interactions per bee species. Tabular view of the information presented in Figure 1 of main text. We used conditional highlighting in the last column (model estimated number of interactions minus raw data) to show which differences were the largest between the two values.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SpeciesID | Number of observed plant interactions per bee species  (from raw GloBI data) | Number of possible plant interactions per bee species  (based on bee-plant phenology data) | Number of model estimated plant interactions per bee species  (Model estimated mean) | Lower 95% CI | Upper 95% CI | Difference in number of plant interactions between model estimate and raw data |
| Andrena angustitarsata | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena anisochlora | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena auricoma | 2 | 122 | 7.67 | 2.05 | 17.25 | 5.67 |
| Andrena baeriae | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena caerulea | 0 | 125 | 0.67 | 0.00 | 1.90 | 0.67 |
| Andrena candida | 0 | 125 | 6.33 | 0.00 | 18.05 | 6.33 |
| Andrena chlorogaster | 0 | 122 | 6.33 | 0.00 | 18.05 | 6.33 |
| Andrena chlorura | 0 | 122 | 0.67 | 0.00 | 1.90 | 0.67 |
| Andrena duboisi | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena gnaphalii | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena hypoleuca | 0 | 125 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena nigripes | 1 | 102 | 3.00 | 1.00 | 6.70 | 2.00 |
| Andrena nigrocaerulea | 1 | 125 | 17.67 | 1.00 | 48.50 | 16.67 |
| Andrena oenotherae | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena pallidifovea | 0 | 122 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena perimelas | 0 | 117 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena piperi | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena plana | 1 | 117 | 3.00 | 1.00 | 6.70 | 2.00 |
| Andrena principalis | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena prunorum | 0 | 122 | 1.33 | 0.00 | 3.80 | 1.33 |
| Andrena sola | 0 | 117 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena suavis | 0 | 122 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena subaustralis | 0 | 117 | 0.33 | 0.00 | 0.95 | 0.33 |
| Andrena submoesta | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Andrena subtilis | 0 | 117 | 0.67 | 0.00 | 1.90 | 0.67 |
| Andrena vandykei | 0 | 124 | 0.00 | 0.00 | 0.00 | 0.00 |
| Pseudopanurgus californicus | 0 | 120 | 0.00 | 0.00 | 0.00 | 0.00 |
| Panurginus armaticeps | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Panurginus gracilis | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Panurginus melanocephalus | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Perdita layiae | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Anthophora californica | 5 | 125 | 16.33 | 5.10 | 35.50 | 11.33 |
| Anthophora edwardsii | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Anthophora urbana | 6 | 125 | 50.67 | 8.30 | 108.05 | 44.67 |
| Bombus californicus | 7 | 125 | 117.67 | 104.10 | 125.00 | 110.67 |
| Bombus crotchii | 7 | 125 | 109.33 | 81.25 | 124.95 | 102.33 |
| Bombus melanopygus | 10 | 125 | 80.67 | 35.55 | 122.95 | 70.67 |
| Bombus occidentalis | 0 | 125 | 48.33 | 1.95 | 118.80 | 48.33 |
| Bombus vosnesenskii | 14 | 125 | 109.33 | 87.60 | 123.70 | 95.33 |
| Ceratina acantha | 2 | 125 | 6.00 | 2.00 | 13.40 | 4.00 |
| Ceratina nanula | 0 | 125 | 0.67 | 0.00 | 1.90 | 0.67 |
| Diadasia australis | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Diadasia bituberculata | 2 | 125 | 42.00 | 3.60 | 87.20 | 40.00 |
| Diadasia nigrifrons | 0 | 117 | 0.33 | 0.00 | 0.95 | 0.33 |
| Diadasia nitidifrons | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Diadasia ochracea | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Diadasia rinconis | 1 | 125 | 6.33 | 1.00 | 16.20 | 5.33 |
| Epeolus minimus | 1 | 123 | 0.00 | 0.00 | 0.00 | -1.00 |
| Eucera actuosa | 0 | 117 | 0.67 | 0.00 | 1.90 | 0.67 |
| Eucera amsinckiae | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Eucera edwardsii | 0 | 117 | 1.00 | 0.00 | 2.85 | 1.00 |
| Eucera frater | 0 | 81 | 0.00 | 0.00 | 0.00 | 0.00 |
| Eucera lunata | 1 | 117 | 10.00 | 1.00 | 26.65 | 9.00 |
| Anthophorula cerei | 0 | 115 | 1.00 | 0.00 | 2.85 | 1.00 |
| Habropoda depressa | 2 | 117 | 39.67 | 2.85 | 94.05 | 37.67 |
| Habropoda miserabilis | 0 | 122 | 13.33 | 0.00 | 38.00 | 13.33 |
| Melecta separata | 0 | 81 | 0.00 | 0.00 | 0.00 | 0.00 |
| Melissodes lupinus | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Melissodes lustrus | 0 | 41 | 0.00 | 0.00 | 0.00 | 0.00 |
| Melissodes stearnsi | 0 | 125 | 0.33 | 0.00 | 0.95 | 0.33 |
| Melissodes tepidus | 0 | 123 | 3.00 | 0.00 | 8.55 | 3.00 |
| Nomada edwardsii | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Nomada formula | 0 | 104 | 0.33 | 0.00 | 0.95 | 0.33 |
| Nomada hesperia | 0 | 122 | 0.33 | 0.00 | 0.95 | 0.33 |
| Nomada semisuavis | 0 | 0 | 0.00 | 0.00 | 0.00 | 0.00 |
| Nomada suavis | 0 | 125 | 2.33 | 0.00 | 6.65 | 2.33 |
| Triepeolus utahensis | 0 | 104 | 0.33 | 0.00 | 0.95 | 0.33 |
| Brachymelecta californica | 1 | 125 | 0.00 | 0.00 | 0.00 | -1.00 |
| Colletes hyalinus | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Colletes kincaidii | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Colletes slevini | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Hylaeus mesillae | 1 | 125 | 0.33 | 0.00 | 0.95 | -0.67 |
| Hylaeus polifolii | 1 | 125 | 0.00 | 0.00 | 0.00 | -1.00 |
| Hylaeus rudbeckiae | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Agapostemon texanus | 16 | 125 | 31.00 | 15.25 | 56.10 | 15.00 |
| Augochlorella pomoniella | 12 | 125 | 32.00 | 12.40 | 61.80 | 20.00 |
| Halictus farinosus | 10 | 125 | 35.67 | 10.65 | 71.45 | 25.67 |
| Halictus harmonius | 0 | 123 | 0.33 | 0.00 | 0.95 | 0.33 |
| Halictus ligatus | 9 | 125 | 17.67 | 8.10 | 33.75 | 8.67 |
| Halictus rubicundus | 0 | 125 | 1.00 | 0.00 | 2.85 | 1.00 |
| Halictus tripartitus | 19 | 125 | 34.67 | 16.30 | 63.80 | 15.67 |
| Lasioglossum albohirtum | 0 | 100 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum argemonis | 0 | 104 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum avalonense | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum brunneiventre | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum cabrilli | 0 | 125 | 4.00 | 0.00 | 11.40 | 4.00 |
| Lasioglossum californiae | 0 | 81 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum channelense | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum diversopunctatum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum hammondi | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum hyalinum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum incompletum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum kincaidii | 0 | 125 | 0.33 | 0.00 | 0.95 | 0.33 |
| Lasioglossum mellipes | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum microlepoides | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum miguelense | 0 | 102 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum nevadense | 0 | 125 | 4.67 | 0.00 | 13.30 | 4.67 |
| Lasioglossum nigrescens | 0 | 96 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum ovaliceps | 1 | 122 | 5.00 | 1.00 | 12.40 | 4.00 |
| Lasioglossum perichlarus | 0 | 115 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum punctatoventre | 0 | 117 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum tegulare | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Lasioglossum titusi | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Anthidium collectum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Anthidium illustre | 0 | 122 | 0.67 | 0.00 | 1.90 | 0.67 |
| Anthidium maculosum | 4 | 125 | 10.33 | 2.05 | 24.85 | 6.33 |
| Ashmeadiella cactorum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ashmeadiella chumashae | 0 | 100 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ashmeadiella cubiceps | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ashmeadiella opuntiae | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ashmeadiella timberlakei | 0 | 125 | 0.33 | 0.00 | 0.95 | 0.33 |
| Ashmeadiella xenomastax | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Atoposmia copelandica | 0 | 81 | 0.00 | 0.00 | 0.00 | 0.00 |
| Coelioxys octodentatus | 0 | 123 | 0.00 | 0.00 | 0.00 | 0.00 |
| Dioxys productus | 0 | 0 | 0.00 | 0.00 | 0.00 | 0.00 |
| Megachile angelarum | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Megachile apicalis | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Megachile brevis | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Megachile coquilletti | 0 | 125 | 0.33 | 0.00 | 0.95 | 0.33 |
| Megachile fidelis | 0 | 123 | 0.67 | 0.00 | 1.90 | 0.67 |
| Megachile montivaga | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Megachile onobrychidis | 0 | 125 | 0.67 | 0.00 | 1.90 | 0.67 |
| Megachile rotundata | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia albolateralis | 0 | 125 | 1.00 | 0.00 | 2.85 | 1.00 |
| Osmia austromaritima | 0 | 81 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia bruneri | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia californica | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia calla | 0 | 125 | 0.67 | 0.00 | 1.90 | 0.67 |
| Osmia clarescens | 0 | 117 | 0.33 | 0.00 | 0.95 | 0.33 |
| Osmia coloradensis | 1 | 125 | 3.33 | 1.00 | 7.65 | 2.33 |
| Osmia gabrielis | 0 | 124 | 1.67 | 0.00 | 4.75 | 1.67 |
| Osmia gaudiosa | 0 | 125 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia kincaidii | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia laeta | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |
| Osmia nemoris | 0 | 125 | 2.33 | 0.00 | 6.65 | 2.33 |
| Osmia tristella | 0 | 125 | 0.67 | 0.00 | 1.90 | 0.67 |
| Protosmia rubifloris | 0 | 122 | 0.00 | 0.00 | 0.00 | 0.00 |