Title
Leveraging Large Biological Interaction Data to Quantify Plant Specialization by Bees

Permalink
https://escholarship.org/uc/item/33b2t2bq

Authors
Bachelder, Nicholas R.
Chen, Angel
Zoe, Fang
et al.

Publication Date
2021-07-01

Supplemental Material
https://escholarship.org/uc/item/33b2t2bq#supplemental

Data Availability
The data associated with this publication are in the supplemental files.

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/
INTRODUCTION

Large, open-access biological datasets, like those hosted by Global Biotic Interactions (GloBI) at https://www.globalbioticinteractions.org/data, have become increasingly accessible due to greater data collection, compilation, and improved storage. These data providers serve to better inform our understanding of species occurrences, interactions, and ecosystem structure. We sought to develop a better understanding of bee specialization of floral resources, an evolutionary trait in bees that underscores the stability and structure of pollinator interaction networks (Fig. 1). GloBI and expert-collected data were compared to better understand patterns in resource specialization. Within our analysis, it became clear that there was a significant correlation between the number of recorded observations and number of unique interactions for any specific bee family, genus, or species.

METHODOLOGY

- Downloaded GloBI bee data (n = 294,914 bee interactions).
- Determined what types of bees were ‘specialists’ or ‘generalists’ using networks and heatmaps.
- Degree of specialization (degree value) is defined as the number of unique plant family interactions a bee species had within the GloBI data.
- Set different cutoff points (no cut, 5 interaction cut) for how many interactions a bee species had to have with a plant family in order to contribute to the degree count.
- Identified an expertly defined list of specialist bees in America authored by Jarrod Fowler (Fowler, 2020).
- Transformed degree value to separate the ‘specialist’ and ‘generalist’ groups for better classification accuracy. Fig. 2 is the formula describing how the transformed degree was generated.
- Trained supervised machine learning models (Decision Trees, Support Vector Machine, Logistic Regression) using the number of recorded observations, degree values, and the expert defined list to create a quantitative classification system to define specialist and non-specialist bees.
- In order to obtain a consistent conclusion, differences in results between various taxonomic levels of bees were analysed.

RESULTS

- Figs. 3 and 4 demonstrate our results including:
  - Strong positive correlation between number of recorded observations and the number of plant families visited.
  - After transforming our degree, modeling achieves above 80% validation accuracy.
  - Binary classification of bee pollinator specialization not consistent

DISCUSSION

We found that while bees expertly classified as bee specialists constantly visited fewer plant families than other bees in the GloBI dataset, there are clusters of species that diverge from the expected trend. These findings indicate that observer bias, on a global scale, can skew our definition of resource specialization or generalization. Moreover, large, open-access datasets like GloBI can change our previous understanding of biological interactions and systems by accessing novel data sources and aggregating.

Although our models weren’t precisely accurate in predicting a bee species as a specialist or generalist, they pave a possible future path of predicting a rank (spectrum) of specialization by using number of observations and number of different plant taxa visited. This degree of classification allows for some biases unlike the binary classifier method we attempted.

CONCLUSION

We found several sources of bias such as how rare bees have fewer recorded observations, making them harder to classify.

We found a strong positive correlation between the number of recorded observations and the number of plant families visited by that same bee species.

Defining specialization of bee pollinators as binary (as generalists or specialists) is not an effective method for describing bee – plant interactions. In observing the supposed bee specialists from the Fowler dataset, the GloBI dataset showed us that these bees interacted with more than just their presumed specialist plant family.

With all of these findings, it is evident that big data can help challenge our assumptions, but also should be carefully utilized to avoid leaving unsolved assumptions unacknowledged.

How can we leverage big data, and in particular biological interaction data, to better understand plant-pollinator specialization? How well defined are the terms we use to describe plant-pollinator specializations?

Total Number of Bee Species: 2504
Total Number of Plant Families: 249

Total Number of Bee Species & Plant Families Within GloBI

Fig. 5 are summary statistics from the bee-plant interactions found in the GloBI dataset including the top 10 bee species and the top 10 plant families recorded. Recorded observations or records in GloBI come from many sources including published literature, observations (including community science) and natural history collection records.

This project was supported by the National Science Foundation projects: Central Coast Data Science Partnership: Training a New Generation of Data Scientists (DSC:1924240) and Digitalizing collections to trace parasite-host associations and predict the spread of vector-borne disease (DBI:1901503).