Metadata for data from: “Habitat restorations in an urban landscape rapidly assemble diverse pollinator communities that persist”

Jens Ulrich, May 2024

This file contains metadata on the five key pieces of data accompanying the manuscript:

- flower\_resources\_herb\_quadrat.csv (the data that was collected on flower counts in the herb enhancements or analogous lawn space)

- flower\_resources\_woody.csv (the data that was collected on flower counts from woody plants)

- land\_cover\_by\_site\_reduced.csv (lat/long and landscape buffer impervious surface and tree canopy cover summaries for each of the 18 park sites)

- pollinator\_data.csv (pollinator detections and the plants that they were recorded interacting with)

- clarkia\_pollination\_data\_2022.csv (seed set for control and supplemented flowers from the pollen limitation experiment)

- reduced\_pollinator\_plant.csv. These are the publicly available interaction data from the following source that were used to supplement our internal interaction data gathered within our study system. We paired down the data set, removing the “Shrub-Steppe” interactions which are from a different biogeographic region. In the R code we also remove interactions south of 48 degree latitude. See the following doi for the original authors data column descriptions. Guzman, L. M., T. Kelly, and E. Elle. 2023. “A Data Set for Pollinator Diversity and Their Interactions with Plants in the Pacific Northwest.” Ecology 104 (2): e3927. <https://doi.org/10.1002/ECY.3927>.

- appendix\_s2.csv - pollinator species information. These data summarize the interaction specialization estimated by our interaction network as well as species-specific phenology as estimated by our model.

**flower\_resources\_herb\_quadrat.csv**

YEAR: year in which data was collected. 1 == 2021, 2 == 2022, 3 == 2023.

SAMPLING\_ROUND: visit within year in which data was collected. Ranges from 1 – 7. Round 1 in year 1 was a practice round that we did not include (we also didn’t have permission to lethally sample pollinators at the beginning of this round). Year 2 we visited some sites a 7th time but did not include these data in our analysis due to difficulties with flexibility of programming around stan’s inability to handle NA values.

DATE: date on which survey was conducted.

SITE: park site name.

SPECIES: plant species recorded.

NUM\_FLORAL\_UNITS: number of open floral units recorded for species across all 20 quadrats for the visit.

**flower\_resources\_woody.csv**

YEAR: year in which data was collected. 1 == 2021, 2 == 2022, 3 == 2023.

SAMPLING\_ROUND: visit within year in which data was collected.

DATE: date on which survey was conducted.

SITE: park site name.

SPECIES: plant species recorded.

NUM\_FLORAL\_UNITS: number of open floral units recorded for species across all 20 quadrats for the visit.

SHRUB\_OR\_TREE: we counted additional plants if we were unsure whether to deem a woody tree/shrub or otherwise. After the surveys were done, the main author (JU) added “y” if the species should indeed be considered a tree/shrub.

**land\_cover\_by\_site\_reduced.csv**

SITE: park site name.

Latitude: latitude of the site centroid.

Longitude: longitude of the site centroid.

Category: herbaceous enhancement or site used as a control.

mean\_herb\_scaled: mean herbaceous enhancement flower abundance across all surveys\*years. Z-score scaled.

mean\_woody\_scaled: mean woody plant flower abundance across all surveys\*years. Z-score scaled.

mean\_herb: mean herbaceous enhancement flower abundance across all surveys\*years. Raw value scaled.

mean\_woody: mean woody plant flower abundance across all surveys\*years. Raw value scaled.

imp\_standardized: amount of area in the buffer that is impervious surface relative to the area of the buffer (which is not always full 500m buffer given that some sites have water bodies in the area and we didn’t want to penalize for this).

canopy\_standardized: amount of area in the buffer that has tree canopy cover relative to the area of the buffer (which is not always full 500m buffer given that some sites have water bodies in the area and we didn’t want to penalize for this).

**pollinator\_data.csv**

YEAR: year in which data was collected. 1 == 2021, 2 == 2022, 3 == 2023.

SAMPLING\_ROUND: visit within year in which data was collected.

DATE: date on which survey was conducted.

SITE: park site name.

CLADE: Syrphidae (hover flies; included in analyses), Anthophila (wild bees; included in analyses), other == we sometimes captured other flies or wasps if we were unsure in the field and then confirmed IDs in the lab and filtered out with R if they were neither of the two focal groups.

SPECIES: pollinator species ID.

COLLECTED\_SPECIMEN: y == we brought back to the lab to ID, physical specimen is associated; n == identified in the field, no physical specimen is associated.

UNIQUE\_SPECIMEN\_ID: unique year/# label given to the pollinator. This is printed on the labels for all physical specimens.

PLANT\_NETTED\_FROM\_FAMILY: Family of the plant that the pollinator was interacting with.

PLANT\_NETTED\_FROM\_GENUS: Genus of the plant that the pollinator was interacting with.

PLANT\_NETTED\_FROM\_SCI\_NAME: Genus species name of the plant that the pollinator was interacting with.

**clarkia\_pollination\_data\_2022.csv**

FLOWER\_UNIQUE\_ID: each flower in the experiment was given a unique ID number.

SITE: the park where the flower was placed.

SITE\_TREATMENT: whether or not the park had an herbaceous enhancement.

POT\_NUMBER: plants were grown seven per pot. We tracked plant groupings by pot.

PLANT\_NUMBER: Within each pot we used 4 plants to for the experiment. Labelling as 1:4 allowed us to group pairs of flowers (control versus supplement) from the same plant.

FLOWER\_TREATMENT: control == ambient pollination; treatment == hand pollinated.

SEEDS\_PRODUCED: Number of seeds produced by the fruiting capsule of the flower.

BOTH\_FLOWERS\_OPEN\_AND\_POLLEN\_APPLIED: Did both flowers open up during the experimental exposure? Did we catch the stigma well receptive for pollen limitation. If yes, good; if no, we couldn’t use the pair for the analysis.

BOTH\_CAPSULES\_RECOVERED: Did both capsules make it back to the common garden and through development without any physical damage? If yes, good; if no, we couldn’t use the pair for the analysis.

PL: 1 – open\_pollinated/hand\_pollinated (defined by Larsen and Barrett 2000 as the number proportion of seeds that the flower could have produced if pollination was optimal)

PL\_INDEX: the complement of above (1 – PL). Our definition describing the proportion of seeds produced relative to the hand pollinated flower.

DATE\_PLACED\_AT\_SITE: Date that the pot was dropped off at a field site.

DATE\_RETURNED: Date that the pot was returned from a field site.

DATE\_FLOWER\_OPEN\_CONFIRMED: we confirmed that a flower was open and receptive at least at some point during the window between the placed and returned dates.

DATE\_POLLEN\_APPLIED: Date at which the first dose of pollen was applied. We applied additional doses if the stigma still appeared receptive on additional site visits. We visited sites 4 total times over the exposure period to apply pollen supplements.

**Appendix\_S2.csv - pollinator species information**

species – the scientific name of the pollinator species

clade – Anthophila indicates bees, Syrphidae indicates hoverflies

total captures – the total number of times a species was collected during the study.

phenology peak – the species specific phenology peak estimated by our model. A value of zero indicates that species detectability (and presumably flight window) peaks at the mean survey date across our visits within years (approximately mid June). A negative value indicates that a species was estimated to reach a peak detection earlier in the season versus a positive value indicating that a species was estimated to reach a peak detection later in the season.

phenology decay – the species specific phenology decay estimated by our model. A value of zero would indicate no change in detectability throughout the season. A negative value indicates that the species becomes less detectable before or after the peak detection date. A more negative value indicates that the detection rate declines more rapidly around the peak detection date (i.e., that the species has a short flight season).

pollen\_specialization – information about oligolecty classification wherever available, taken from <https://jarrodfowler.com/pollen_specialist.html>. We didn’t use this in our models.

The following 6 metrics are calculated using only our internal interaction data from Vancouver city parks.

degree – the number of plant species that a pollinator species was recorded interacting with.

normalised\_degree – the number of plant species that a pollinator species was recorded interacting with divided by the total number of plant genera interacted with by any pollinator species in our dataset.

d – the species specific specialization metric Bluthgen’s d (*d*’) as estimated using the bipartite package in R. Calculated after grouping plants by genus.

degree\_scaled – z-score scaled degree.

normalised\_degree\_scaled – z-score scaled normalised\_degree

d\_scaled – z-score scaled d

The above 6 metrics were recalculated with a “supplemented” tag at the end of the column name (e.g., “degree\_supplemented”) using our internal interaction data from Vancouver city parks combined with the extrernal data from the publicly available dataset on pollinator interactions from our region (see metadata list of files for that dataset).

The above 6 metrics were then recalculated with a “supplemented\_genus” tag at the end of the column name (e.g., “degree\_supplemented”) using our internal interaction data from Vancouver city parks combined with the extrernal data from the publicly available dataset on pollinator interactions from our region (see metadata list of files for that dataset) BUT with plants grouped by genus before the metrics were calculated. These are the metrics we used in our final analyses presented in the manuscript. We grouped by genus to resolve some disagreement or uncertainty in species level identifications across the datasets and also to avoid overinflating generalization based on interactions with many closely related species.