2024

Do Generalist-Pollinated Plants Exhibit Pollination Syndromes? Examination of Floral Size, Shape, Color, and Scent on Pollinator Services to Wild Sunflowers (Helianthus)

Charles A. Pitsenberger II
University of Central Florida

DO GENERALIST-POLLINATED PLANTS EXHIBIT POLLINATION SYNDROMES? EXAMINATION OF FLORAL SIZE, SHAPE, COLOR, AND SCENT ON POLLINATOR SERVICES TO WILD SUNFLOWERS (*HELIANTHUS*)

By

CHARLES ALBERT PITSENBERGER II
B.S., University of Alabama, 2020

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Biology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2024

Major Professor: Chase Mason

ABSTRACT

Traditional pollination syndromes group angiosperms into categories based on how floral traits impact the functional group of pollinator most associated with those traits. The concept, while well supported for specialist-pollinated plants, is a poor predictor of pollinator identity in generalist systems, such as those common to the family Asteraceae. One potential avenue for future refinement of the concept is the combination of large floral trait datasets, quantitative pollinator data, and phylogenetic comparative methods. *Helianthus* is a well-studied genus of North American aster whose species include the agriculturally significant *H. annuus*, which represents the third largest oilseed crop globally. The genus is primarily bee pollinated and, while much is known about traits that are correlated with bee attraction at short ranges common to agricultural and horticultural settings, there has been little research on long range visual and chemical attraction traits within the genus. Using data on display size and shape, ray color, floral volatile composition, and floret depth collected from *Helianthus* species grown in a common garden, mixed models were constructed to predict pollinator visitation as a function of floral traits. For four of seven pollinator response variables, there was at least one model that outperformed null models, and three of the four best models were multivariate. This work will inform future research of pollination syndromes within generalist systems such as those common to Asteraceae.

I dedicate this thesis to the bees.

ACKNOWLEDGMENTS

I would like to thank everyone who helped with data collection, analysis, and who supported me throughout this project. Including but not limited to:

All of the undergrads that I've had the pleasure to work with: Colin Fitz, Alina Makarenko, Aleyna Harris, Samantha Plock, Janna Ferdaus, Sara Zecca, Antonia Garcia, Priya Seth, Corinne McComb, Marie Preuss, & Renato Guzman.

My brilliant, humble, phenomenally talented lab-mates: Rebekah Davis, Roberta Beard, Jenny Bouchenot, and Keivan Bahmani.

My committee: Chase Mason, Patrick Bohlen, and Eric Goolsby.

And finally, Jesse Nash, without whom I would not have developed the skills in Excel required to process data quickly and efficiently.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AR1- Anthocyanin Reflectance Index 1 (Gitelston et al. 2009)

B- Brightness, the sum total of light in the visible portion of a reflectance spectrum

C- Chroma, the saturation of color

CR1- Carotenoid Reflectance Index 1 (Gitelston et al. 2009)

ELPD- Expected log predictive density

FRI- Flavonol Reflectance Index (Merzylak et al. 2005)

GC-MS Gas Chromatography-Mass Spectrometry

GLMM- Generalized linear mixed model

H- Hue, a quantification of the shade or color of a ray

HERA- Horticultural and Ecological Research Area

LM- Long-medium contrast

MCMC- Markov-chain Monte Carlo chain

MS- Medium-short contrast

PCA- Principal component analysis

PGLS- Phylogenetic Generalized Least Squares regression

SRc- Simple ratio index for carotenoids (Manjunath et al. 2016)

UV- Ultraviolet

W- The first quartile of the bee visible light spectrum, from 300-400nm

X- The second quartile of the bee visible light spectrum, from 400-500nm

Y- The third quartile of the bee visible light spectrum, from 500-600nm

Z- The fourth quartile of the bee visible light spectrum, from 600-700nm

CHAPTER ONE: INTRODUCTION

The evolution of the flower and subsequent adaptations for animal pollination mark an important development in the evolution of angiosperms (Specht & Bartlett 2009). Using conspicuous displays, angiosperms attract pollinators that can carry their pollen directly to other flowers, allowing them to reproduce more efficiently and over a greater range than is possible with wind and other abiotic means of pollination. Christian Konrad Sprengel (1793) was the first to formally document the phenomenon of animal pollination, and Charles Darwin (1877) was the first to link the evolution of floral traits to selective pressure from pollinators. Since the publication of these seminal works, we have advanced our knowledge of plant-pollinator interactions and have dubbed the collection of traits that result from selective pressure from pollinators to be a plant's "pollination syndrome" (Fenster et al. 2004). The traditional pollination syndrome groups floral traits (and the species that possess them) into syndromes based on the pollinator functional group that the traits are associated with, e.g. red flowers are commonly associated with the hummingbird pollination syndrome (Fenster et al. 2004). This traditional concept is not without controversy. Theoretically one could infer a plant's primary pollinator by merely observing a flower and noting traits that are indicative of certain pollinating taxa, but several studies have found that this is often not the case (Dellinger 2020; Ollerton et al. 2009; Fenster et al. 2004). A review of recent literature on pollination syndromes identifies one potential avenue for future refinement of the concept - the combination of large floral trait datasets, quantitative pollinator data, and phylogenetic comparative methods (Dellinger 2020). The same review also notes that there is relatively little work on pollination syndromes within the most diverse plant family: Asteraceae.

Nestled within Asteraceae, the genus *Helianthus* (sunflowers) contains some 50 or so species of North American wildflower (Stephens et al. 2015), including the agriculturally significant *Helianthus annuus* (Mason et al. 2017). The genus is well-studied, with a plethora of morphological, physiological,

and genetic data widely available, as well as a well resolved phylogeny of the diploid members of the genus (Mason et al. 2017; Bahmani et al. 2022; Stephens et al. 2015; Todesco et al. 2022). The genus shows a high degree of phenotypic diversity and is distributed throughout a range of habitats, including forests, wetlands, prairies, deserts, and beaches (Mason et al. 2017). The flat yellow capitula of *Helianthus* species indicates a bee pollination syndrome, and multiple studies corroborate this relationship in both wild and crop sunflowers (Rogers 1988, Posey et al. 1986). Additionally, several traits such as nectar content, floret size, pollen, and ultraviolet (UV) pigmentation have been studied for their effect on pollinator visitation in *H. annuus* (Todesco et al. 2022; Ferguson et al. 2021; Portlas et al. 2018; Mallinger et al. 2017; Chabert et al. 2022). These traits are understood to influence foraging choices of bees at short ranges seen in agricultural settings, but less is understood of traits related to long range attraction, particularly visual and chemical cues.

Using a combination of approaches meant to capture a large proportion of variation in floral phenotype, I here present a study of traits that contribute to the pollination syndrome of *Helianthus* species. The body of research on *Helianthus*, shared pollinators, and genetic and phenotypic diversity within the genus make it a system that is well-suited for such a study. A preliminary analysis of morphological and volatile data of 16 species of *Helianthus*, combining data from two studies of plants grown at different times and locations (Mason et al. 2017; Bahmani et al. 2023), found substantial correlations between certain morphological and volatile traits. I hypothesize that relationships between otherwise unrelated traits are resultant of a shared pollination syndrome within *Helianthus*, and that statistical models incorporating said traits will successfully predict visitation rates of bees to *Helianthus* capitula. To address these hypotheses, I employ phylogenetic comparative methods to build models that predict pollinator visitation rates as a function of floral traits.

CHAPTER TWO: METHODS & RESULTS

Experimental Methods

Study Area

As wild *Helianthus* species occupy geographically and biotically diverse habitats with highly variable climate and soil conditions, a traditional common garden approach was employed to assess species under common growing conditions and a shared pollinator community. The Horticultural and Ecological Research Area (HERA) garden, established in the summer of 2022, consists of two hundred total plots of *Helianthus* species spread across two replicate gardens with spatial randomization (Figure 1). Seeds for 100 different accessions spanning 49 species were sourced from the United States Department of Agriculture National Plant Germplasm System (USDA NPGS) via the Germplasm Resource Information Network (GRIN). For each species, a minimum of two accessions were chosen to represent populations from different regions within each species' range, capturing both intraspecific and interspecific trait variation. For species with multiple subspecies, one accession was chosen per subspecies.

Plants were germinated in a uniform manner. Seeds were scarified in the laboratory and placed on wet filter paper in Petri dishes in the dark until roots emerged. Once root hairs were present, seedlings were moved under growth lights with a 12-hour photoperiod until cotyledons became green. Seedlings were then transplanted into seedling trays containing moist sand. After developing multiple pairs of true leaves, young plants were transferred to 6-inch pots filled with potting soil in a greenhouse setting, and then transplanted into their assigned HERA garden plots once large enough to have a high rate of successful outdoor establishment. A target of 6-12 individual plants were assigned to each 4 m^2 plot, all consisting of the same accession (wild source population). As the diverse annual and perennial species did not germinate at the same rate, re-germination and transplantation efforts continued from

summer 2022 through spring 2023 with a focus on the least successfully established accessions to fill in space in plots to achieve the target plant density. For perennials, establishment in 2022 permitted vigorous growth in the second year (during data collection for this study), while annual species either self-seeded within plots or were re-germinated and planted in spring 2023. Once plants were present in the HERA garden, all plots were treated with the same watering regime, with supplemental watering provided daily for the first few days post-transplant, and weekly thereafter when precipitation was absent. . As is often the case for common gardens, not all plants thrived, due to either chance events like deer browsing or mismatch between native habitats and local conditions. Of the 100 target accessions that were selected for inclusion, the majority reached flowering, and 53 accessions are included in this study representing 31 different wild *Helianthus* species, or ~62% of species within the genus.

Figure 1: A bumblebee in the North HERA garden, foraging on *H. debilis* subsp. *tardiflorus* while an observer collects pollinators from *H. argophyllus.* Photographed by Jenna Palmisano.

Traits Assessed

The floral traits assessed in this study were selected to capture a range of pollinator-relevant aspects of capitula diversity - visual display traits like size and pigmentation, chemical traits relating to volatile fragrance, and handling traits based on disc floret size (Table A1). To assess these traits for each accession, the sampling unit for this study is the individual plot – representing one spatial replicate of one accession of a given species. With two duplicate gardens and two accessions per species, each species could be sampled up to four times. For each plot, a minimum of three capitula were sampled, and plot-level trait averages were used for subsequent data analysis. Sampling took place approximately three times per week from mid-May to mid-November 2023, spanning over six months. For sampling, live capitula were removed from plants in the field, placed into vases of water, and brought to the laboratory for processing. Capitula were removed at the same time period on sampling days, in the mid-morning from the hours of 9:00 am-11:00 am local time. Sampling was standardized by choosing the largest capitula with the most open disc florets from each plot. Floral trait data was collected from least to most destructive: first standardized photographs were taken, then ray florets were plucked for hyperspectral reflectance measurements, then disc florets were collected for volatile analysis via solid phase microextraction gas chromatography mass spectrometry (SPME-GC-MS), and finally digital scans of extracted disc florets were taken.

Standardized Photographs

Standardized photographs were used to capture the morphological diversity of *Helianthus* (Figure 2), and to obtain total display area, ray length, ray area, disc area, disc:ray ratio, ray number, and ray density (Table A1). Capitula were placed into a light box containing a frame with a scale consisting of 4 dots each 13mm in diameter forming a square of 40cm per side. Capitula were photographed from a distance of ~1m by an 18 Mp Canon Rebel T6 DSLR camera (Canon, Inc; Tokyo, Japan). A tag was included in each photograph identifying the sampling unit for each capitulum, and three capitula from

each sampling unit were photographed. Photographs were analyzed using ImageJ (Schneider et al. 2012). After setting the scale, disc and total capitulum diameter were measured from three angles, and used to calculate averages for total display area, ray area, disc area, and disc:ray ratio using the formula $A = \pi r^2$. Ray length was calculated similarly, as an average of the lengths of three rays. Ray numbers were taken as a raw count, and ray density were calculated as ray number divided by disc circumference.

Figure 2: Diversity of *Helianthus* species photographed in this study. From left to right, top to bottom: *H. argophyllus, H. carnosus, H. verticillatus, H. debilis* subsp. *cucumerifolius, H. resinosus, H. maximiliani, H. decapetalus, H. praecox* subsp. *praecox, H. radula.*

Ray Hyperspectral Reflectance

Hyperspectral reflectance of ray florets was measured using a CI-710 SpectraVue Leaf Spectrometer (CID Bioscience, Inc; Camas, WA), collecting reflectance, absorbance, and transmittance from 300-1100nm from ray florets plucked from three capitula from each sampling unit. This data was used to quantify ray hue, brightness, and chroma, as well as derive estimates of ray anthocyanin, carotenoid, and flavonol content using the methods detailed below. Because it is known that concentration of pigments (including those responsible for ultraviolet patterning; Todesco et al. 2022) vary along the length of sunflower rays, measurements were taken at the base, middle, and tip of each ray to capture this variation. In cases where rays were too small to fully cover the sensor, multiple rays were used and arranged such that the same part of the ray (i.e. base middle and tip) covered the sensor. In the case of particularly short rays, as determined by the amount of overlap between potential base middle and tip measurements, measurements were truncated to include only base and tip or, sometimes, one total measurement. This occurred with relative infrequency, and was primarily restricted to *H. debilis, H. porteri,* and *H. microcephalus* which have the smallest rays in the genus. In instances of only two measurements, variables derived from hyperspectral data were averaged to estimate the "middle" value. For the very smallest rays where only one measurement could be taken, measured values were considered representative of base, middle, and tip measurements and were entered three times into the final dataset. Ray color was quantified using segment classification of reflectance spectra based on chroma, hue, and brightness (Smith 2014; Endler 1990). Simple reflectance indices consistent with Gitelson et al. (2009) were used to estimate anthocyanin and carotenoid content, and flavonol content was estimated similarly (Merzlyak et al. 2005).

Per Endler (1990) and Muchhala et al. (2014), overall brightness (B), or the sum total of visible light reflecting from the ray, is calculated as the area under the curve for the visible light portion of a

reflectance spectrum (R(λ)) (eq 1). For the purposes of this study, the bee visible light spectrum, from 300-700nm, will be used (Muchhala et al. 2014).

$$B = \int_{300}^{700} R(\lambda) d\lambda \quad (1)$$

Chroma and hue are calculated by first dividing the overall spectrum into four discrete quartiles, representing theoretical color ranges in the bee visible spectrum: W(300-400nm), X(400-500nm), Y(500-600nm), and Z(600-700nm). Absolute brightness, B_W, B_X, B_Y, and B_Z are calculated as the area under the reflectance curve for each range (eq 2a-2d):

$$B_W = \int_{300}^{400} R(\lambda) d\lambda \quad (2a)$$

$$B_X = \int_{400}^{500} R(\lambda) d\lambda \quad (2b)$$

$$B_Y = \int_{500}^{600} R(\lambda) d\lambda \quad (2c)$$

$$B_Z = \int_{600}^{700} R(\lambda) d\lambda \quad (2d)$$

Absolute brightness is divided by overall brightness to find relative brightness (eq 3a-3d).

$$W = \frac{B_W}{B} \quad (3a)$$

$$X = \frac{B_X}{B} \quad (3b)$$

$$Y = \frac{B_Y}{B} \quad (3c)$$

$$Z = \frac{B_Z}{B} \quad (3d)$$

To account for the way animal visual receptors function, relative brightness is then converted in terms of contrast between long (L), medium (M), and short (S) wavelengths (eq 4a, 4b), where:

$$LM = Z - X \tag{4a}$$

$$MS = Y - W \tag{4b}$$

Using the values for LM and MS, Chroma (C), or the saturation of color, and hue (H), or the type of color, can be visualized graphically where LM is the value for Y, and MS is the value for X (Smith 2014; Endler 1990). In this representation, chroma is the distance from 0,0 of the point (MS, LM) and is calculated as:

$$C = \sqrt{LM^2 + MS^2} \tag{5}$$

Hue is represented as the counterclockwise angle from y=0 of the point (MS, LM), and is calculated in one of several ways, depending on the value of LM (Smith 2014):

$$if\ LM > 0, H = \left|arcSin(\frac{MS}{C})\right| 2\pi \tag{6a}$$

$$if\ LM < 0, H = \pi - arcSin(\frac{MS}{C}) \tag{6b}$$

In addition to the preceding quantification of floral color, relative concentrations of pigments responsible for floral color were also calculated. For anthocyanins and carotenoids, relative pigment content was estimated using multiple established reflectance indices (Manjunath et al. 2016; Gitelson et al. 2009). The presented equations, 7a for carotenoids and 7b for anthocyanins, represent the indices that were retained and used in analysis. Similarly, equation 7c was used to estimate ray flavonol content (Merzlyak et al. 2005).

$$CR1 = \frac{1}{R_{510}} - \frac{1}{R_{550}} \tag{7a}$$

$$AR1 = \frac{1}{R_{550}} - \frac{1}{R_{700}} \tag{7b}$$

$$FRI = \left({}^{1}/_{R_{400}} - {}^{1}/_{R_{460}}\right) R_{800} \tag{7c}$$

GC-MS

Solid Phase Microextraction Gas Chromatography-Mass Spectrometry (SPME-GC-MS) was used to quantify floral volatiles. For each sampling unit, three 10ml glass headspace vials were loaded with 0.2000 ± 0.0200 g of fresh disc florets. Each vial contained samples from different capitula, such that each vial contained biologically distinct samples. Newly open florets were preferentially sampled. In the case of discs which were too small to obtain sufficient material to fill one vial, multiple discs were used, but material from one disc was never placed into multiple vials. For GC-MS analysis, the protocol previously used by Bahmani et al. (2022) and Anandappa et al. (2023) for analysis of sunflower fragrance was employed. In brief, samples were loaded into a single quadrupole GC-MS-QP2020 (Shimadzu, Inc.; Kyoto, Japan). Vials were incubated for 20 minutes at 90°C and agitated at 250rpm. A 50/30 µm divinylbenzene/carboxen/polydimethylsiloxane (DVS/CAR/PDMS) SPME fiber extracted volatiles from the vial headspace for 10 minutes. The fiber was desorbed in the GC-MS inlet for 3 minutes at 225°C. The fiber was conditioned at 270°C for 10 minutes between samples. Column flow was 1.5ml/min with a split ratio of 25:1, and a purge flow of 3.0ml/min after 4 minutes of sampling time. Initial temperature was set to 30°C for one minute, then increased to 150°C at 12°C/min, held for one minute, increased again to 225°C at 9°C/min, then to 250°C at 50°C/min, and held for 5 minutes. The source temperature for the mass spectrometer was kept at 200°C, and the interface temperature was kept at 250°C. Mass spectra were compared against the National Institute of Standards and Technology (NIST) standards database and peak identities with >70% similarity were selected and used for preliminary compound identities (Lemmon et al. 2017). Spectra were aligned and a combination of similarity hits from the mass spectra and retention time was used to manually validate compound identities. In instances where preliminary identity and retention time did not line up with other spectra, alternative hits with comparable similarity were checked. If no alternative hits were found with retention times that match up with more trustworthy peak identities from other spectra in the dataset, then the peak was not considered reliable

enough to include in this study. For each compound, peak area was divided by the sample's mass to calculate mass normalized peak area, and compounds were manually classified by chemical group (i.e. monoterpene, diterpene, sesquiterpene, phenolic, etc.). Total volatile abundance, monoterpene proportion, diterpene proportion, and sesquiterpene proportion were calculated from the peak areas for compounds in each sample. Relative abundance of compounds that were present in >100 samples was calculated as the percent composition of each compound, per-sample.

Corolla Scans

Corolla depth was measured using a CanoScan LiDE flatbed scanner (Canon, Inc; Toyko, Japan). Direct measurement of the corolla itself is difficult, so the entire floret length was used as a reliable substitute as previously reported for sunflower (Portlas et al. 2018). Scans were taken at a density of 300 dots per linear inch (DPI), meaning that each scan has a known standardized scale. The same tags used in the standardized photographs, which themselves have a standard size, were included in each scan as both an indicator of sampling unit, as well as a secondary scale bar. From three capitula per sampling unit, five disc florets were plucked with forceps and arranged on the scanner. The nearest open florets to the disc's center were preferentially sampled to standardize selection and to represent florets that were being actively foraged by pollinators at the time of sampling. ImageJ was used to measure average floret length for each set of 15 florets per sampling unit.

Pollinator Visitation Surveys

Pollinator visitation was assessed via plot-level visual surveys. Pollinator surveys were conducted on each flowering plot containing ≥10 capitula throughout its flowering period. Taking place from June-October 2023, a total of 176 visual surveys were performed, and 4,777 pollinator visits were recorded. Surveys were conducted in the morning, primarily between the hours of 8am and 11am local time. The start and end time was recorded for each survey day, so as to calculate the approximate time of each

survey. Surveys were conducted cyclically, such that each flowering plot was surveyed before any were repeated. For each survey, a plot's total capitulum count was taken, so as to correct for its effect on visitation. For plots with particularly large displays, a subsample of capitula was surveyed, and the total capitula count as well as the number surveyed was recorded. Otherwise, the plot was observed in its entirety. For five minutes per survey, a plot was monitored and each visit by a potential pollinator was recorded, as well as the taxonomic identity of the visitor to the best of the observer's ability. To reduce bias, all surveys were carried out by the same observer. The end result of each survey was a count of capitula that were surveyed, the total capitula present in the plot, and a tally of visitors from each group. For the purposes of this study, a "visitor" is defined as any flying insect who contacts a disc, as even an incidental landing may result in the transport of pollen and thus may exert a nonzero selective pressure on the plant. Visits from organisms other than insects were not recorded, for instance spiders who are unlikely to move between capitula with any regularity. A capitulum being visited by multiple insects at once was counted as a visit for each insect, and one insect moving from disc to disc was counted as a visit for each disc on which it landed. For real-time identification purposes, taxonomic identities were grouped into the following categories: *Bombus, Apis, Xylocopa,* Megachilidae, Halictidae*,* Other bee, Other Hymenoptera, Diptera, Lepidoptera, and Other Insect. These categories represent a combination of the most specific group that the observer could identify accurately in the field, and the least specific group deemed to be useful to this study. For example, more attention is paid to bees as they are by far the most frequent visitors, and because their distinct feeding guilds are known to influence their ability to collect from a given capitulum (Cariveau et al. 2016, Ferguson et al. 2021, Portlas et al. 2018). In the case of visitors falling into "other" categories, or into Diptera or Lepidoptera, an effort was made to identify each visitor to at least the Order level.

For each survey, the sum of all visits was taken. The traits used as response variables for subsequent modeling are as follows: total pollinator count per survey, per-survey count of bees from

each of the following three taxonomic groups: *Bombus*, Halictidae, and Megachilidae, and per-survey proportion of visitors from each of the three aforementioned taxonomic groups.

Covariates

Environmental conditions are known to influence pollinator abundance and activity (Corbet et al. 1993). To account for this across the six-month survey period, the following data was recorded for use as covariates in modeling: the date of survey (as the ordinal day of year), the approximate time of survey (as minutes after midnight, approximated based on known survey duration, order, and day start and end times), the per-plot capitulum count per survey, the total precipitation on the day of the survey, and the temperature at the time of survey. . Per-plot capitulum count was measured as both the number of capitula actively surveyed, as well as the total number of capitula in the plot at time of survey. To attain temperature and precipitation data, historical data was pulled using the WeatherSTEM Data Mining Tool (WeatherSTEM, 2024) for the FSWN University of Central Florida weather station, the closest station to the site of the HERA garden (~1 mile).

Data Curation

Data was first curated by data type (GC-MS, reflectance, etc.) before being incorporated into the larger dataset(s). Standardized photos were hand measured in ImageJ. Each photograph was processed by at least two individuals, as a form of internal validation. Since reflectance data was measured along the same wavelengths, the spectra were naturally aligned when combined into one spreadsheet. Data was not recorded in whole number increments of wavelengths so, for wavelength values in formulas 1-7, the nearest measured wavelength to each given integer was identified and used in its place. The trapezoidal rule of calculus was employed to estimate the area under the curve for the purposes of equations 1-6. All calculations were done in Microsoft Excel (v. 2405; Microsoft Corporation), using the values of reflectance at wavelengths from formulas 1-7, and the integration for reflectance spectra, as

appropriate. GC-MS data was compiled onto one spreadsheet, aligned, and compound IDs were manually validated by retention time. Compounds present in more than 100 samples were selected to be treated as their variables for subsequent analysis, and their per-sample percent composition was recorded. To ensure accuracy, particular attention was paid to these compounds and compounds at their retention times during manual validation. Siloxanes and other non-plant compounds known to be artifacts of the GC column were removed, and peak area was mass normalized. After validation, a cutoff criterion was implemented to prune the dataset of peaks that were present in fewer than three samples and made up less than 1% of total peak area. Compounds that were not removed were manually classified as either monoterpene, diterpene, sesquiterpene, or other.

Once curated, individual datasets were combined into one master document. Individual trait measurements were averaged across repetitions for plot-level data, which was averaged across gardens for accession-level data and averaged across species for species-level data. Species-level data was used to identify phylogenetic correlations between traits. To produce the dataset used for modeling of pollinator visitation, plot-level trait data as well as data on covariates was added to the visitation dataset.

Analysis

Phylogenetic Comparative Methods

To determine the relationship between predictor and response variables in a system where all species are closely related, it is necessary to use an approach capable of parsing out the effect of relatedness (i.e., the hierarchical nonindependence of species) to show how consistent the relationship is across the phylogeny. To account for the effects of relatedness on trait distributions, I used a variety of methods that incorporate phylogeny into the linear models and correlation matrices used in this study. R version 4.3.1 was used for all analyses (R Core Team 2023). The *ape* (Paradis et al. 2019) and *Rphylopars* (Goolsby et al. 2024) packages were used to generate correlation matrices for use in identifying potential

pollination syndromes. For modeling, the *brms* (Bürkner, 2017) package was used. For all phylogenetic comparative methods, an adapted version of the most recent *Helianthus* phylogeny (Stephens et al. 2015) was used. Species not assessed within the present study pruned from the tree, and taxa not present in the original phylogeny were grafted in with placement based on best evidence across available genetic and taxonomic sources.

Pollination Syndromes in *Helianthus*

To examine whether pollination syndromes exist within *Helianthus*, evidence of correlations among the different floral traits was assessed. If seemingly unrelated floral traits correlate strongly with one another across sampling units, such as disc area and monoterpene proportion, then this may be preliminary evidence of a pollination syndrome. In R, an initial principal components analysis (PCA) was performed, which found no significant axes of covariation, with the first principal component only explaining 26% of variation in the dataset. Subsequently, pairwise correlation matrices were used to examine whether individual traits might be forming pollination syndromes not captured by an overall PCA. Before conducting phylogenetically-explicit analyses, Pearson correlations were used to generate a matrix of floral trait variables and used to identify groups of traits that were synonymous ($|r| \geq 0.9$). The list of variables was trimmed down from 100 to 32 distinct, non-synonymous variables. Variables that were removed include, but are not limited to, ray area (synonymous with total display area), disc diameter (synonymous with disc area), and most reflectance variables that were measured at the base and tip. It was identified that measurements at the middle were synonymous with base and tip measurements, but that correlations between base and tip were not strong enough to be flagged though still correlated. Given this, the decision was made to use middle ray measurements as the representative reflectance variables for the ray as a whole. The exception to this was FRI, as it is known that the UV pattern of sunflowers is variable. Thus, FRI for base and tip were retained. Additionally, while multiple indices were calculated for carotenoids and anthocyanins, one index for each was chosen and retained

for use in analysis. CR1 and AR1 (Gitelson et al. 2009) were retained, as they had low correlation with other reflectance traits, and were calculated in a parallel manner.

For evidence of pollination syndromes, a new correlation matrix was made in R using *Rphylopars* (Goolsby et al. 2024). From the full set of 100 floral trait variables, Phylogenetic Generalized Least Squares (PGLS) regression was used on every possible trait pair to generate linear models predicting one variable as a function of another. From each model, the covariance coefficient was extracted and used to calculate correlation between the two traits. The outputs were used to generate a phylogenetically-corrected correlation matrix (Figure 3). From the matrix, suites of traits were identified where each trait was moderately correlated ($|r| \geq 0.5$) with one another. In cases where the potential syndrome contained a variable that was excluded from the non-synonymous set of 32 traits, the variable was replaced with the corresponding synonymous variable in the syndrome. This happened with some frequency, as there were several instances where traits with $|r| \geq 0.5$ were not retained in the non-synonymous dataset, but the trait that was retained had $|r| < 0.5$. For example, the simple ratio for carotenoids (SRc) was moderately correlated with monoterpene proportion, diterpene proportion, and sesquiterpene proportion. It was excluded from the non-synonymous dataset due to its correlation with ray brightness ($r=0.905$), so brightness was incorporated into a model with monoterpene, diterpene, and sesquiterpene proportions (Table A3). The resultant method uncovered a short list of four potential syndromes (Table A3). The same method was then used, with a threshold of $r \geq 0.4$, to uncover five more potential syndromes.

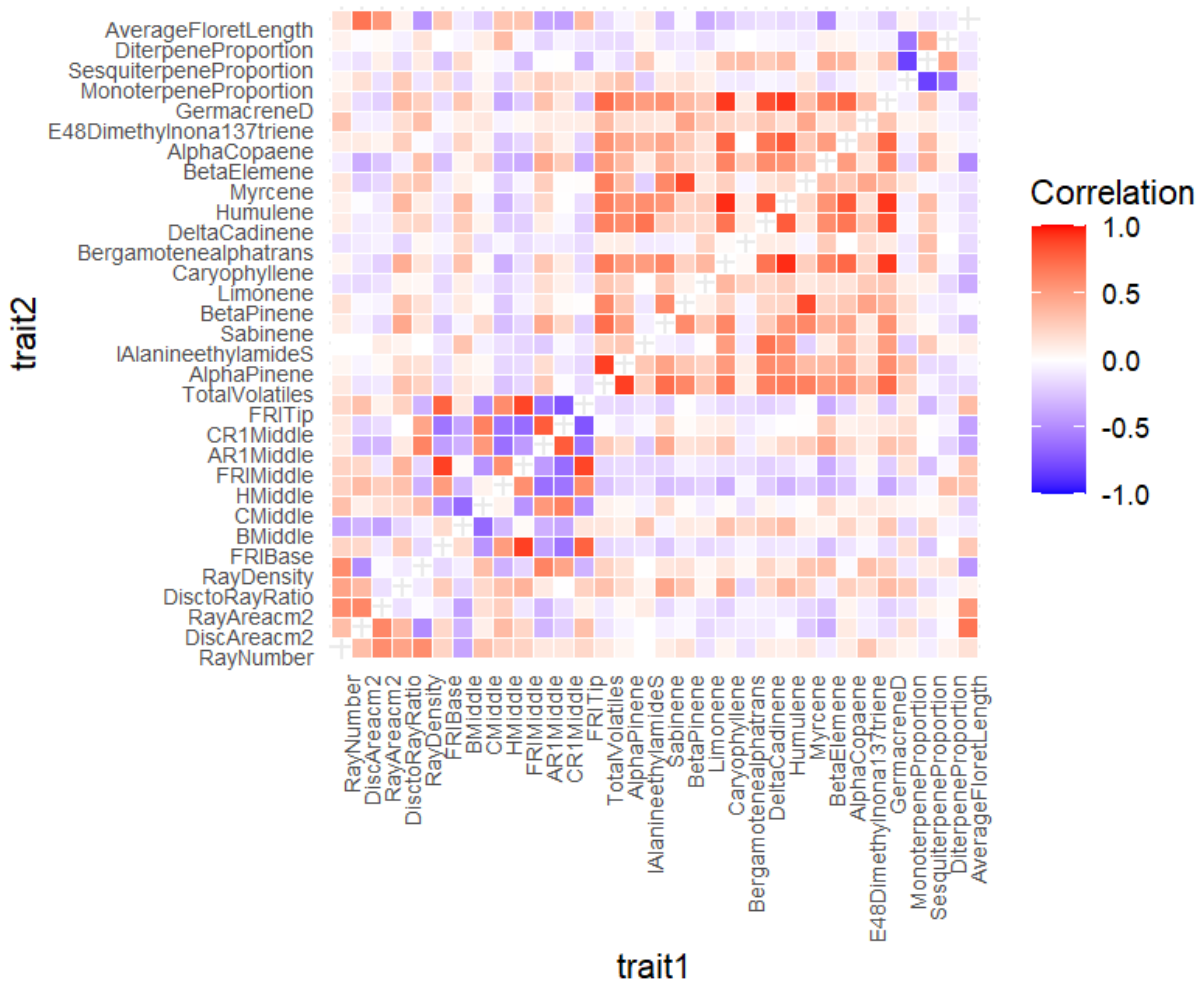

Figure 3: Matrix of Pearson correlations for the 32 traits used as predictor variables, calculated from the covariance coefficient of pairwise PGLS models.

Modeling Floral Trait Impacts on Pollinator Visitations

The floral traits of interest in this study are as follows: *total display area, ray length, ray area, disc area, disc:ray ratio, ray number, ray density, ray hue, ray brightness, ray chroma, ray anthocyanin content, ray carotenoid content, ray flavonol content, total volatile abundance, monoterpene proportion, sesquiterpene proportion, diterpene proportion, monoterpene:sesquiterpene ratio, floret length,* and *percent composition of select volatiles* (Table A1). These were incorporated into linear models as predictor variables. The primary response variables are total pollinator count per survey, per-survey count of bees from each of the following three taxa: *Bombus*, Halictidae, and Megachilidae, and per-

survey proportion of visitors from each of the aforementioned taxa. Potential covariates that were accounted for are date of survey, approximate time of survey, temperature at time of survey, total precipitation on the day of the survey, and per-plot capitula count per survey.

Bayesian generalized linear mixed models (GLMMs) were fit using the *brms* package (Bürkner 2017) to assess associations between pollinator visitation (the response variable) and one or multiple floral traits (fixed effects). For count-based response variables (*Bombus*, Megachilidae, Halictidae, and total pollinator counts), a negative binomial distribution was assumed, while additionally accounting for zero-inflation in the three taxon-specific count-based models. For proportion-based response variables (*Bombus*, Megachilidae, and Halictidae), a zero-one-inflated beta-binomial distribution was assumed.

Each model included the following covariates as fixed effects: the date of survey (ordinal day of year), temperature at time of survey, total precipitation on the day of the survey, the per-plot capitulum count per survey, as well as a block effect corresponding to the plot's location (north or south garden). To account for potential phylogenetic nonindependence in the data, phylogenetic covariance was included as a random effect, and plant accession was included as a random effect to account for variation within species. Additionally, baseline reference models that included only the covariates and random effects were fit to determine the additional predictive power of floral traits.

For all models, four Markov-chain Monte Carlo (MCMC) chains were sampled over 10,000 iterations each, including a warmup of 4,000 iterations, and the *adapt_delta* parameter was set to 0.99 to minimize divergent transitions. All continuous fixed effects were centered and scaled to a mean of 0.0 and a standard deviation of 1.0 to improve MCMC convergence.

To assess the relative performance among candidate models, group *k*-fold cross-validation (*k*=10) was performed, where observations were assigned to folds based on species identity (Roberts et al. 2017; Vehtari et al. 2023). The model with the highest expected log predictive density (ELPD) was

considered the best-supported model, and models with an ELPD within 4 points of the best-supported model (i.e., ELPD difference ≥ -4.0) were considered to be equally supported (Sivula et al. 2020). For any reference models with an ELPD difference ≥ -4.0, floral traits were assumed to have too weak of an effect to be detected as predictive of the corresponding response variable. For each response variable, posterior means and 95% credible intervals for the best-supported models were examined for fixed effects, coefficients of determination (R^2), and *k*-fold R^2.

Results

Phenotypic Variation Among Taxa

Most traits were highly variable, exhibiting threefold to tenfold variation, with some traits such as total volatiles and relative α-pinene abundance showing 100-fold and 1,000 fold-variation, respectively (Appendix, Table A2). There was particularly low variation among several metrics of color, most notably brightness and chroma. After correcting for phylogeny, there was a notable dearth of the expected strong correlations between otherwise unrelated traits. Between unrelated traits, AR1 was moderately correlated with ray density, and floret length was moderately correlated with both disc area and ray area. There were, however, several significant correlations were between traits of the same type, e.g. a negative correlation of -0.82 between sesquiterpene proportion and monoterpene proportion.

Potential Suites of Traits

In total, the first selection criteria (|r| >0.5) found four potential syndromes, with the second, lower finding six more (Table A3). Some minor deviations were made from the rule that all traits be strongly correlated with one another, particularly among syndromes identified using the first selection criteria. Notable deviations were Brightness in the first syndrome, where other traits were sufficiently correlated with one, but not both, of B tip/ B base. In the second syndrome, all traits were correlated

with AR2, but not necessarily each other. And in the third syndrome, SRc was only correlated with diterpene proportion, but not monoterpene or sesquiterpene proportion.

Models

For Megachilidae count, *Bombus* proportion, and Megachilidae proportion, floral traits did not improve predictive power over the null models (|ELPD_diff| > 4). For total pollinator count and *Bombus* count, a single best model was identified. For Halictidae count there were three best models and for Halictidae proportion, there were five models that both improved predictive power over the null models and had |ELPD_diff| < 4.

The best model for total pollinator count (R^2=0.103, kfold R^2=0.042) includes disc and ray area, ray brightness and hue, anthocyanin content, and floret length (Table A4). Within this model, higher visitation was predicted by larger disc area, smaller ray area, and lower brightness and anthocyanin content. Among environmental covariates, total pollinator visitation increased with ambient temperature and the number of capitula sampled per plot, and garden was a had a non-zero slope indicating that the North garden experienced more visits on average. Higher precipitation was associated with fewer visitors, but this term had a non-zero estimate only in the best model and not in the null . Floret length, ray number, ray hue, and date of survey were retained in the best model, but slope estimate credible intervals contained zero indicating no appreciable impact.

The best model for *Bombus* count (R^2=0.10, kfold R^2=0.038) includes ray density, anthocyanin content, and relative quantities of germacrene D and β-elemene (Table A5). In this model, higher *Bombus* visitation was predicted by lower anthocyanin content and lower percent abundance for β-elemene. Higher precipitation was also associated with lower visitation. While retained in the best model, slope estimate credible intervals for ray density, germacrene D, ambient temperature, date of survey, garden and number of capitula sampled all contained zero indicating no appreciable effect.

For Halictidae count, the best model (R^2=0.054, kfold R^2=0.024) includes floret length as the sole predictive floral trait (Table A6). Among the equivalent best models, floret length was the also sole floral trait variable that consistently had a non-zero slope estimate, with shorter florets being associated with more visits. This association between floret length is in opposition to trends seen in other taxa (Figure 4) but is in line with literature on bee feeding guilds, which indicate that bees preferentially forage on corollas that match their tongue lengths (Ferguson et al. 2021, Portlas et al. 2018). Similarly, garden was the only environmental covariate with a consistently non-zero slope estimate, with the North garden experiencing higher visitation. Temperature was found to be significant in only one model, including nulls, and weakly indicated that visitation increased as temperature increased. While retained in the equivalent best models, ray area, disc area, ray number, and sesquiterpene proportion were not found to have an appreciable effect on visitation. Among environmental covariates, date of survey, precipitation, and number of capitula sampled were variously retained in the best models but had no appreciable effect on visitation.

For Halictidae proportion, the best model (R^2=0.037, kfold R^2=0.0219) includes disc and ray area, ray brightness, and floret length (Table A7). Among the five equivalent best models, two models indicated that smaller discs and shorter florets were predictive of a higher proportion of visits by Halictidae relative to other pollinators. Along with these traits, the very best model also indicated that higher ray number was associated with higher Halictidae proportion. Conversely, the other three equivalent best models did not include disc area or floret length, but instead retained anthocyanin content, with higher anthocyanin content associated with a higher proportion of visits. Ray area, ray brightness, ray number, ray hue, ray density, total volatile abundance and percent abundance of germacrene D & β-elemene were retained in various models but did not have non-zero slope estimates. Among covariates, date of survey had a non-zero slope estimate in all models, with earlier dates being associated with a higher proportion of Halictid visitors. Garden had a non-zero slope estimate in all but

one equivalent best model, with a higher proportion of Halictid visits in the North garden. The number of capitula sampled had a non-zero impact in the two models that included disc area and floret length, while temperature and precipitation had no appreciable effects in any of the equivalent best models.

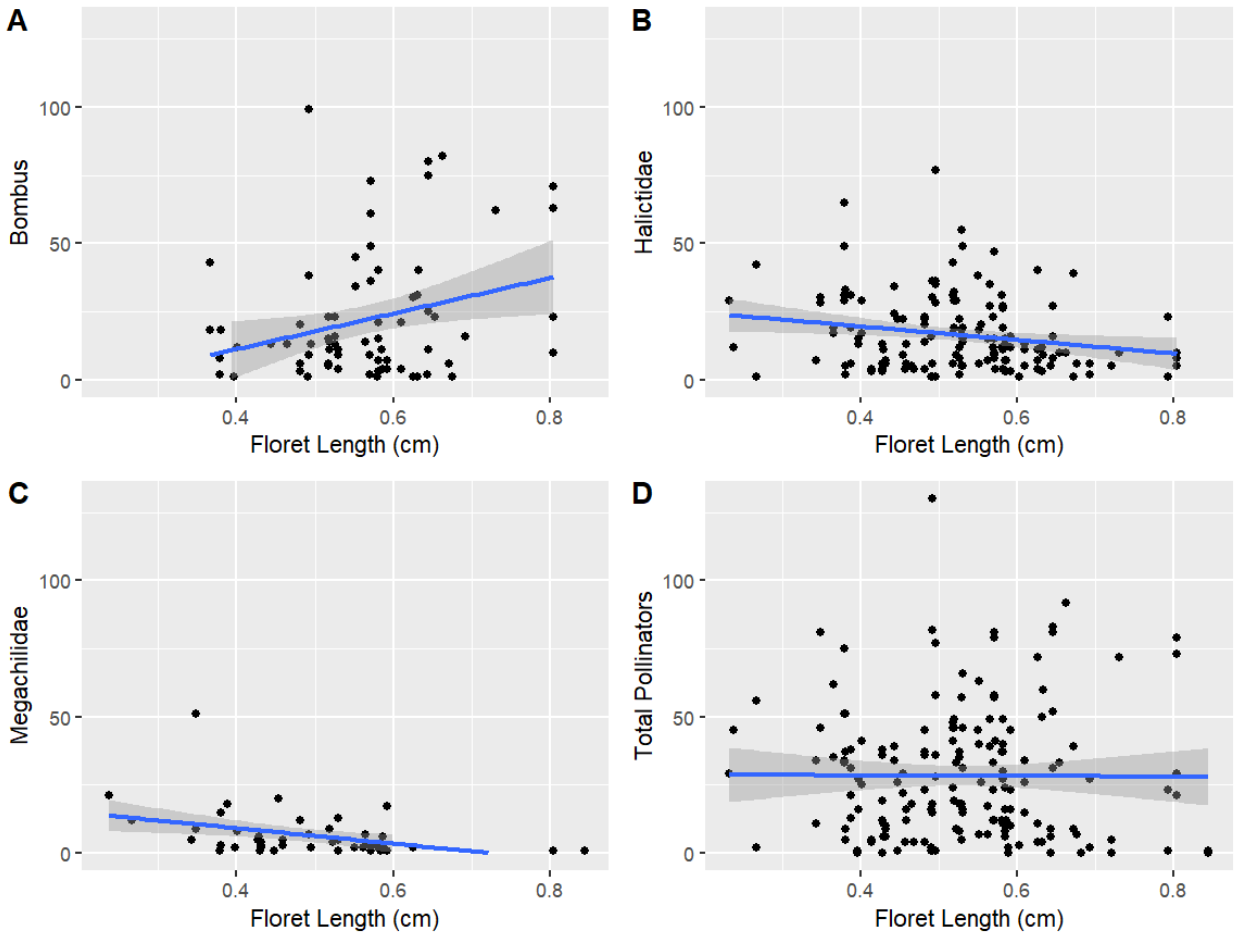

Figure 4: Scatter plots showing trends between floret length and counts of the three main clades studied, as well as total pollinators.

CHAPTER THREE: DISCUSSION

Trait Diversity and Putative Syndromes

An initial PCA found no significant axes of variation. However, the PGLS based correlation matrix did uncover several groups of moderately correlated traits. The lack of significant axes of variation indicates that sunflower inflorescences occupy diverse regions in multivariate floral trait space, meaning that floral traits evolve highly independently in *Helianthus*. Mason et al. (2017) demonstrated that some, but not all, floral traits are significantly correlated with abiotic soil and environmental factors, with large capitula correlated with arid environments and small capitula correlated with wetter environments. The results of this study corroborate the finding that floral trait evolution is indeed complicated in *Helianthus.* Floral traits have likely not evolved in isolation and are subject to many competing selective pressures. However, several moderately correlated groups of traits were identified in this study. Future research is warranted to assess the role of pollinators in the co-occurrence of these traits, as well as the combined effects of biotic and abiotic factors on floral trait evolution in *Helianthus*.

Seasonality and Floret Length

Depending on the type of pollinator assessed, several floral trait variables and environmental covariates had differing effects. Date and floret length, in particular, had numerous differences in magnitude and directionality of the effect. While there were more overall pollinators as the seasons progressed from summer to autumn, the effect of seasonality differed from group to group. This is not unexpected, as it is well known that bee community composition exhibits substantial seasonal variation (Oertli et al. 2005). The effect of floret length differed among groups as well, in a similar manner (Figure 4). The lack of a clear pattern between floret length and total pollinator count is expected, since total visits include both the short- and long-tongued bees, as well as other non-bee pollinators that will have differing relationships to floret size – where insect proboscis length determines access to nectar rewards

(Ferguson et al. 2021; Portlas et al. 2018). *Bombus*, as a long-tongued genus, is expected to prefer long disc florets, as this group can access nectar rewards deep inside the long corolla tube and larger florets typically produce more nectar per floret (Cariveau et al. 2016; Ferguson et al. 2021; Portlas et al. 2018). However, Halictidae, as a short-tongued clade, not only had a negative relationship with floret length, but floret length was the sole trait that was a consistently strong contributor to models predicting Halictid count. This is very likely due to the physical restrictions inherent to being a short-tongued bee, as although long-tongued bees are known to prefer longer florets even though they can forage on all floret sizes, short-tongued bees have no choice but to forage on florets that are of an accessible size (Ferguson et al. 2021; Portlas et al. 2018). Megachilidae, another long-tongued group, was expected to have a positive relationship with floret length as well. The lack of floral trait predictive models for this group is likely due to the lower volume of data available for this group relative to *Bombus* and Halictidae.

Display Morphology

The three variables of disc area, ray area, and ray number were each retained as predictors with non-zero slope estimates in best models for total pollinator count, Halictidae count, and Halictidae proportion, though presence and directionality varied among the different response variables. For total pollinators, larger discs were associated with more visits, while larger ray area was associated with fewer. The relationship among disc area, ray area, and total area is such that ray area correlates significantly with total display area by PGLS ($r = 0.9989$), while disc area does not. The positive relationship between disc area and visitation is likely explained by the additional resources present in larger discs – both nectar and pollen rewards. Conversely, it is possible that negative relationship between ray area and visitation is driven by overall smaller capitula which, due to their limited resources, promote a higher rate of movement of pollinators from capitula to capitula. It is well known that disc size has direct correlations with seed yield in cultivated sunflowers, and beyond the direct 1:1 relationship between floret number and ovule number it may be that this relationship is in part due to the relative

attractiveness of larger discs to pollinators improving realized seed set (Marinković 1992). Perhaps more interesting is the fact that smaller discs lead to a greater proportion of Halictid visitors. Since this relationship only holds true for Halictidae proportion, this may be a case of Halictid bees being overrepresented in surveys of less popular species, due to their overall high presence in the garden. However, disc area does have a relatively strong positive correlation with floret length (Figure 3). Since floret length is such a strong predictor of Halictid visitation, it is entirely possible that the associations between Halictidae proportion and disc area are merely a reflection of the tendency of these short-tongued bees to forage on species with shorter florets.

Color and Bee Vision

Considering ray pigmentation, anthocyanin content (AR1) and brightness were the only variables that were present with non-zero slope estimates in the best models for any pollinator response variable. Brightness, when present with a nonzero slope estimate, had a weakly negative relationship, where lower brightness indicated more visitors. Brightness did have negative correlations with the untested variables of long-medium (LM) and medium-short (MS) contrast (eq. 4a, 4b). Since brightness is a measure of the sum total of light reflected by the ray, and not a measure of color (Hue) or saturation (Chroma), the directionality of the relationships between pollinator visitation and ray brightness may not be a case of bees choosing floral displays that are less bright. Rather, it may be a case of bees choosing displays with more visual contrast, which just so happen to score lower than average for overall brightness.

When AR1 was retained in equivalent best models, its slope was consistently non-zero. However, the directionality of the relationship between anthocyanin content and visitation differed among groups. For total pollinators and *Bombus*, the relationship was negative. However, for Halictidae proportion the association was positive. Since the association was only positive for Halictidae proportion, this suggests

that Halictids do not specifically strongly favor high anthocyanins, but by virtue of being most common visitor, Halictid bees may simply be overrepresented in surveys of plants that are less popular. This may suggest that ray anthocyanin content is selected against by most *Helianthus* pollinators, and indeed human-visible ray anthocyanin content is uncommon across the genus (though frequent in discs, as well as the bracts of some species). Alternatively, small-statured Halictid bees may be forced to choose foraging sites based on characteristics that are generally unpopular among other bees. More research is needed to fully parse out the differences in responses among groups.

Chroma, as a measure of contrast, was anticipated to be an important factor explaining pollinator visitation. However, it is absent from most of the best models, and where present exhibits weak slopes with credible intervals that contain zero. This result may be driven by several different factors. First, of the floral traits chroma had by far the least variation observed. Although ray color is variable among sunflowers to the human eye (Figure 2), it may not be so variable to bees. Plots of long-medium (LM) and medium-short (MS) contrast (Figure 5) were used to visualize both hue and chroma in both human color space as well as that of bees (Muchhala et al. 2014). Though plotted as cartesian coordinates along the axes of LM and MS, each point also represents hue and chroma as polar coordinates, where chroma is R and hue is Θ. When calculated using the different color spectra, the shape and orientation of the data changes drastically. In the human visible light spectrum, the data distribution is a smear with a consistent (yellow) hue of variable saturation (chroma). However, in the bee visible light spectrum, the data distribution is more clustered. This suggests that the aspects of ray color that bees can detect in sunflowers are less variable than our own eyes would indicate. This is an intriguing finding and suggests that ray color has diversified far less than other floral traits as *Helianthus* has adapted to habitats across North America – a pattern consistent with stabilizing selection. Future work should consider explicit tests of how sunflower floral color variation is perceived by bees, and how this may alter visitation with other traits held constant. For example, the ornamental cultivated

sunflower germplasm contains a much larger variety of ray color than present in wild *Helianthus*, with relatively less variation in morphological traits. Cultivated sunflower may be a suitable system to address this question, along with specific known color variants of wild sunflower (e.g., white color morphs of *Helianthus debilis*).

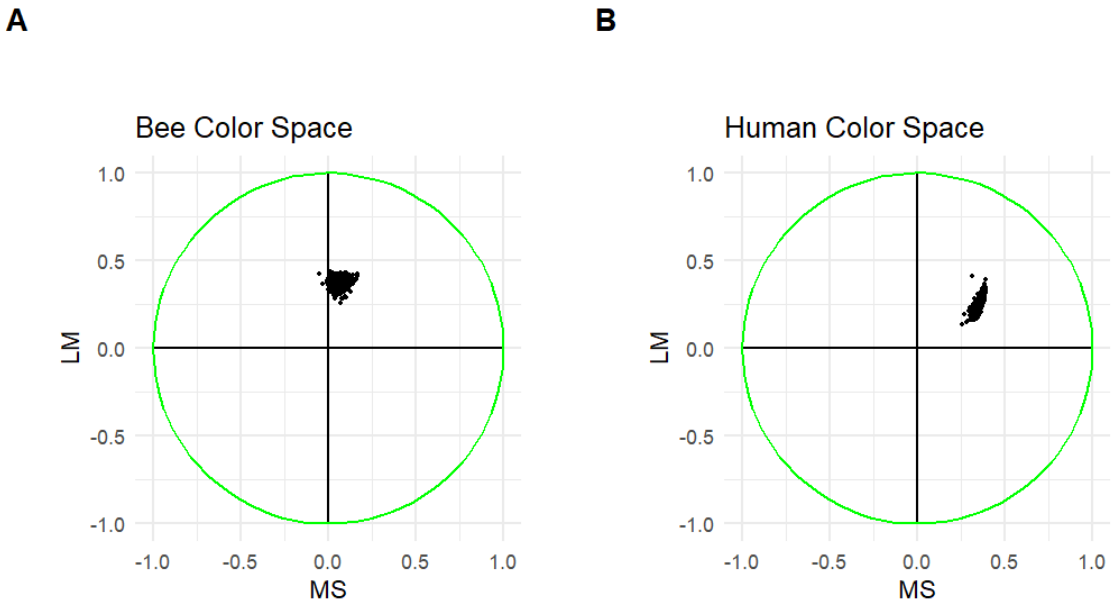

Figure 5: Visualization of ray chroma (C) and hue (H) for ray florets, calculated using segment classification for the human visible spectrum (400-700nm), and for the bee visible spectrum (300-700nm).

Volatiles

Whenever they were retained with non-zero slopes in many of the equivalent best models for pollinator responses, variables related to the abundance of floral volatiles affected visitation negatively across the board. Furthermore, the primary volatile produced by sunflowers, α-pinene, was entirely absent from any of the best models. This was entirely unexpected, as scent is known to be a factor that influences pollinator foraging decisions and, indeed, speciation within many specialized systems (Benelli et al. 2017; Schiestl 2014). The original expectation was that higher-fragrance species would be more

attractive, with scent detectable by pollinators from a farther distance. However, in generalist and bee-pollinated systems, it has been found that volatile profiles are often dominated by a single compound, as is the case for *Helianthus* with α-pinene. In these systems, it is thought that volatiles serve as a form of associative learning, where pollinators cue into the presence of a specific scent, rather than a complex mix of scents (Schiestl 2014). If this is the case, then it would stand to reason that the primary volatile produced by sunflowers, α-pinene, would not necessarily affect visitation directly but serve as a reliable cue that the capitula belong to *Helianthus*. In this scenario, an abundance of other "adulterants" such as Germacrene D or β-elemene would negatively affect visitation by altering the associative cue relative to the typical profile of sunflowers within the garden as a whole. Additionally, β-elemene has known insecticidal properties, possibly explaining its apparent repelling properties with regards to bumblebees (Govindarajan & Benelli 2016). This suggests that bee attraction by floral fragrance is much more complex than expected, but also leaves open the question of why *Helianthus* floral profiles vary so much in composition beyond the dominant α-pinene and sabinene compounds, with over a hundred compounds detected at >1% of mass-normalized peak area in this study.

Sources of Error

Although four of seven pollinator response variables had models that improved predictive power compared to the null model, the R^2 values for models were low across the board. This could be due to a variety of factors. It is possible that there are one or more untested variables that influence pollinator attraction, such as pollen or nectar availability. If these traits are variable across *Helianthus* and not significantly correlated with any of the predictor variables tested, then the low R^2 of models could be due to the fact that the traits tested truly don't have a strong effect on pollinator visitation. It is also possible that the effects of the response variables were being clouded by the effects of one or more covariates. Temperature and seasonality are known to have strong influences on pollinator activity, but the two are strongly correlated with one another (Oertli et al. 2005). It is possible that a subsequent

careful re-analysis using fewer covariates, or structured analyses like directed acyclic graphs or structural equation modeling, may be able to more appropriately disentangle the impacts of ambient conditions relative to variation in floral traits. Likewise, a sliding-window analysis explicitly considering the flowering phenology of the focal species within this dataset may permit additional insights into relative foraging preferences of the seasonally available pollinator community in choosing among the available *Helianthus* species. Overall, pollinator visitation data is highly variable and the present statistical analysis, while valid and rigorous, is not necessarily the most optimized approach to such complex time-series data. Subsequent optimization may show stronger predictive power than present results would indicate.

Conclusions

Among the best predictive models for visitation, all but two (Table A6, Model 1; Table A7, model 4) were multivariate, stemming from the list of potential pollination syndromes that were compiled based on correlations between traits. For three of seven pollinator variables tested, the best predictive model stemmed from the list of potential pollination syndromes. Due to the relatively low R^2 values for models across the board, and the relatively low number of multivariate models tested, it is difficult to call the results of this study anything more than suggestive. However, for total pollinators, *Bombus* count, and proportional Halictidae models, the primary hypothesis that suites of traits identified via phylogenetically corrected correlations will form the best predictive models for pollinator visitation was supported. For Halictidae count, *Bombus* proportion, Megachilidae count, and Megachilidae proportion, the primary hypothesis was not supported, as the best model for Halictidae count contained a single floral trait and there were no models that outperformed the nulls for Megachilidae count, Megachilidae proportion, or *Bombus* proportion.

Floral traits were found to be highly variable, with no significant axis of variation. Although floral traits have likely evolved independently in *Helianthus,* it is still entirely possible that trait combinations

co-occur with relative frequency, and that said co-occurrence is related to specific pollinator groups. Future research should focus on the weak floral trait syndromes identified in this study, and should aim to shed more light on the differential effects that differences in morphology, color, and scent have on differential visitation by the various groups of pollinators, particularly long-tongued (e.g., *Bombus* and Megachilidae) versus short-tongued bee groups (e.g., Halictidae). Trends between predictor and response variables for Halictidae compared to other pollinators were frequently inverted, and future work is necessary to determine whether this can be explained by floret length alone, or whether there truly is a suite of floral traits that short-tongued bees alone prefer in *Helianthus*.

APPENDIX A: SUPPLEMENTARY TABLES

Table A1: Traits which were considered as predictor, response, and covariates.

Response Variables	***Predictor Variables***	***Covariates***
Total pollinator count	Total capitulum area	Time of day
Bombus count	Ray length	Date
Halictidae count	Disc area	Maximum daily Temperature
Megachilidae count	Ray area	Precipitation
Bombus proportion	Disc:ray ratio	Average daily temperature
Halictidae proportion	Ray number	per-plot capitulum count
Megachilidae proportion	Ray density	
	Floret length	
	Ray hue (base, middle, tip)	
	Ray brightness (base, middle, tip)	
	Ray chroma (base, middle, tip)	
	Ray Anthocyanin content (base, middle, tip)	
	Ray Carotenoid content (base, middle, tip)	
	Ray Flavonol content (base, middle, tip)	
	Total volatile abundance	
	Monoterpene proportion	
	Sesquiterpene proportion	
	Diterpene proportion	
	% δ-Cadinene abundance	
	% α-pinene abundance	
	% Humulene abundance	
	% L alanine ethylamide abundance	

Response Variables	*Predictor Variables*	*Covariates*
	% Myrcene abundance % Sabinene abundance % β-Eleme abundance % β-Pinene abundance % α-Copaene abundance % Limonene abundance % (E)-4,8-Dimethylnona-1,3,7-triene abundance % Caryophyllene abundance % Germacrene D abundance % Trans-α-bergamotene abundance	

Table A2: Species-level means, standard deviations, minima, and maxima of floral traits used as predictor variables.

Trait	Mean	Standard Deviation	Minimum	Maximum
Ray Number	12.46 cm	4.77cm	0cm	23.83cm
Disc Area	2.15 cm^2	1.73 cm^2	0.31 cm^2	9.51 cm^2
Ray Area	27.91 cm^2	16.75 cm^2	0 cm^2	78.94 cm^2
Disc to Ray Ratio	16.31	7.52	0.024	0.15
Ray Density	2.72	0.80	0	4.58
Flavonol Reflective Index (Base)	-9.72	2.27	-15.08	-5.21
Ray Brightness (Middle)	1.34E+04	1.23E+03	1.08E+04	1.74E+04
Ray Chroma (Middle)	0.39	0.021	0.32	0.42
Ray Hue (Middle)	1.19	0.38	0.47	2.06
Flavonol Reflective Index (Middle)	-9.17	2.00	-13.68	-5.45
Anthocyanin Reflective Index 1 (Middle)	0.0027	0.00086	0.0010	0.0046
Carotenoid Reflective Index 1 (Middle)	0.080	0.041	0.014	0.18
Flavonol Reflective Index (Tip)	-8.47	1.89	-13.45	-5.29
Total Volatiles	4.09E+07	3.44E+07	1.18E+06	1.50E+08
α-Pinene	2.65E+07	2.20E+07	2.81E+05	1.05E+08
l Alanine ethylamide (S)	2.48E+05	2.02E+05	0	9.73E+05
Sabinene	1.44E+06	1.99E+06	0	1.14E+07
β-Pinene	3.08E+06	6.47E+06	0	5.25E+07
Limonene	1.09E+06	1.30E+06	0	5.48E+06
Caryophyllene	2.96E+05	5.12E+05	0	2.88E+06
Bergamotene (α-, trans-)	3.19E+05	4.65E+05	0	2.36E+06
δ-Cadinene	1.35E+05	1.55E+05	0	1.08E+06
Humulene	1.06E+05	1.55E+05	0	8.58E+05
Myrcene	1.27E+06	2.32E+06	0	1.57E+07
β-Elemene	1.30E+05	2.12E+05	0	1.05E+06
α-Copaene	4.69E+04	6.99E+04	0	2.41E+05
(E)-4,8-Dimethylnona-1,3,7-triene	1.37E+05	2.11E+05	0	1.25E+06
Germacrene D	7.54E+05	1.40E+06	0	7.54E+06
Monoterpene Proportion	0.82	0.18	0.14	1.00
Sesquiterpene Proportion	0.093	0.94	0	0.42
Diterpene Proportion	0.0066	0.012	0	0.070
Floret Length	0.53 cm	0.12cm	0.23cm	0.84cm

Table A3: Potential pollination syndromes that were identified, as well as any substitutions that were made and rationale.

Initial Trait Suite	**Substitutions**	**Final Potential Pollination Syndrome**	**Rationale**
Disc area, ray area, *B base, B tip*, floret length, ray number	B middle for B base & B tip	Disc area, ray area, B middle, floret length, ray number	|r| >0.5
Ray Density, *AR2*, *Total sesquiterpenes,* β-elemene	AR1 for AR2, Germacrene D for total sesquiterpenes	Ray density, AR1, Germacrene D, β-elemene	|r| >0.5
Total volatiles, *total sesquiterpenes, total diterpenes*, α-pinene, sabinene, β-pinene, caryophyllene, δ-cadinene, humulene, myrcene, β-elemene, α-copaene, germacrene D	Total sesquiterpenes and total diterpenes removed with no substitute	Total volatiles, α-pinene, sabinene, β-pinene, caryophyllene, δ-cadinene, humulene, myrcene, β-elemene, α-copaene, germacrene D	|r| >0.5
Monoterpene proportion, diterpene proportion, sesquiterpene proportion, *SRc middle*	SRc for B middle	Monoterpene proportion, diterpene proportion, sesquiterpene proportion, B middle	|r| >0.5
Disc area, ray area, *B base, B tip*, floret length, ray number, H middle, AR1	B middle for B base and B tip	Disc area, ray area, B middle, floret length, ray number, H middle, AR1 middle	|r| >0.4
Disc: ray ratio, sabinene, caryophyllene	none	Disc: ray ratio, sabinene, caryophyllene	|r| >0.4
B middle, *total sesquiterpenes*, H middle, *AR2*	AR1 for AR2, Germacrene D for total sesquiterpenes	B middle, Germacrene D, H middle, AR1	|r| >0.4
AR1, total volatiles, *total sesquiterpenes*	Germacrene D for total sesquiterpenes	AR1, total volatiles, Germacrene D	|r| >0.4
Floret length, sesquiterpene proportion	none	Floret length, sesquiterpene proportion	|r| >0.4

Table A4: Slope estimates and 95% credible intervals for the highest performing models as well as null models for total pollinators, R^2, kfold R^2, and ELPD difference for each model relative to the best model. Slopes with credible intervals that do not contain zero are bolded.

Total Pollinators	*Model 1*	*Environmental covariate Null*	*Intercept only Null*
Disc Area	**0.56 (0.19, 0.93)**	-	-
Ray Area	**-0.51 (-0.87, -0.17)**	-	-
B middle	**-0.27 (-0.50, -0.05)**	-	-
Floret Length	-0.10 (-0.35, 0.15)	-	-
Ray Number	-0.11 (-0.37, 0.14)	-	-
H middle	-0.27 (-0.56, 0.01)	-	-
AR1	**-0.41 (-0.73, -0.09)**	-	-
Julian Date	0.21 (0.00, 0.42)	0.17 (-0.04, 0.39)	-
Temperature	**0.44 (0.24, 0.65)**	**0.39 (0.18, 0.60)**	-
Precipitation	**-0.18 (-0.33, -0.02)**	-0.12 (-0.27, 0.04)	-
Number Sampled	**0.23 (0.04, 0.42)**	0.17 (-0.01, 0.35)	-
Garden	**-0.30 (-0.58, -0.02)**	**-0.29 (-0.57, -0.01)**	-
R2	0.103	0.082	0.041
kfold R2	0.0422	0.0099	0.0072
ELPD_diff	0	-9.83	-13.63

Table A5: Slope estimates and 95% credible intervals for the highest performing models as well as null models for *Bombus* count, R^2, kfold R^2, and ELPD difference for each model relative to the best model. Slopes with credible intervals that do not contain zero are bolded.

Bombus	*Model 1*	*Environmental covariate Null*	*Intercept only Null*
Ray Density	-0.23 (-0.68, 0.24)	-	-
AR1	**-0.89 (-1.38, -0.42)**	-	-
Germacrene D	0.30(-0.15, 0.76)	-	-
ß-Elemene	**-1.12 (-1.87, -0.47)**	-	-
Julian Date	0.15 (-0.44, 0.73)	-0.07 (-0.70, 0.51)	-
Temperature	0.48 (-0.03, 1.01)	0.31 (-0.24, 0.88)	-
Precipitation	**-0.47 (-0.94, -0.02)**	**-0.48 (-0.96, -0.018)**	-
Number Sampled	0.38 (-0.07, 0.90)	0.43 (-0.08, 0.97)	-
Garden	-0.03 (-0.66, 0.6390)	-0.006 (-0.68, 0.69)	-
R2	0.1	0.092	*0.076*
kfold R2	0.038	0.0066	*0.0091*
ELPD_diff	0	-20.4	*-19.07*

Table A6: Slope estimates and 95% credible intervals for the highest performing models as well as null models for Halictidae count, R^2, kfold R^2, and ELPD difference for each model relative to the best model. Slopes with credible intervals that do not contain zero are bolded.

Halictidae	*Model 1*	*Model 2*	*Model 3*	*Environmental covariate Null*	*Intercept only Null*
Ray Area	-	-0.27 (-0.63, 0.06)	-	-	-
B Middle	-	-0.13 (-0.31, 0.05)	-	-	-
Floret Length	**-0.34 (-0.51, -0.16)**	**-0.35 (-0.59, -0.12)**	**-0.33 (-0.51, -0.15)**	-	-
Disc Area	-	0.24 (-0.10, 0.60)	-	-	-
Ray Number	-	0.09 (-0.13, 0.30)	-	-	-
Sesquiterpene Proportion	-	-	-0.05 (-0.24, 0.13)	-	-
Julian Date	-0.15 (-0.38, 0.09)	-0.124 (-0.36, 0.12)	-0.13 (-0.37, 0.11)	-0.10 (-0.33, 0.14)	-
Temperature	0.17 (-0.05, 0.38)	**0.22 (0.01, 0.43)**	0.17 (-0.05, 0.37)	0.18 (-0.04, 0.40)	-
Precipitation	-0.04 (-0.20, 0.12)	-0.08 (-0.25, 0.09)	-0.04 (-0.21, 0.12)	-0.04 (-0.21, 0.13)	-
Number Sampled	-0.04 (-0.23, 0.15)	0.03 (-0.17, 0.23)	-0.04 (-0.23, 0.15)	0.07 (-0.13, 0.27)	-
Garden	**-0.41 (-0.68, -0.12)**	**-0.41 (-0.70, -0.13)**	**-0.40 (-0.68, -0.12)**	**-0.41 (-0.72, -0.11)**	-
R2	0.054	0.074	0.056	0.035	0.023
kfold R2	0.024	0.031	0.024	0.0079	0.0077
ELPD_diff	0	-0.6	-0.9	-8	-10.6

Table A7: Slope estimates and 95% credible intervals for the highest performing models as well as null models for Halictidae proportion, R^2, kfold R^2, and ELPD difference for each model relative to the best model. Slopes with credible intervals that do not contain zero are bolded.

Halictidae proportion	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Environmental covariate Null*	*Intercept only Null*
Disc Area	**-0.73 (-1.27, -0.23)**	**-0.67 (-1.25, -0.16)**	-	-	-	-	-
Ray Area	0.45 (-0.05, 1.02)	0.47 (-0.03, 1.06)	-	-	-	-	-
B Middle	-0.070 (-0.32, 0.17)	0.06 (-0.25, 0.37)	-	-	-	-	-
Floret Length	**-0.53 (-0.92, -0.19)**	**-0.45 (-0.86, -0.10)**	-	-	-	-	-
Ray Number	**0.40 (0.03, 0.74)**	0.32 (-0.06, 0.68)	-	-	-	-	-
H middle	-	0.16 (-0.23, 0.58)	-	-	-	-	-
AR1	-	0.33 (-0.11, 0.77)	**0.36 (0.09, 0.62)**	**0.45 (0.19, 0.69)**	**0.40 (0.13, 0.65)**	-	-
Ray Density	-		0.16 (-0.13, 0.44)	-		-	-
Germacrene D	-	-	0.12 (-0.16, 0.40)	-	0.06 (-0.25, 0.36)	-	-
ϐ-Elemene	-	-	0.30 (-0.01, 0.63)	-		-	-
Total Volatiles	-	-		-	0.19 (-0.08, 0.49)	-	-
Julian Date	**-0.42 (-0.74, -0.12)**	**-0.45 (-0.77, -0.14)**	**-0.48 (-0.79, -0.16)**	**-0.37 (-0.69, -0.07)**	**-0.39 (-0.73, -0.06)**	**-0.39 (-0.72, -0.06)**	-
Temperature	-0.10 (-0.41, 0.20)	-0.11 (-0.42, 0.20)	0.02 (-0.29, 0.34)	-0.08 (-0.38, 0.22)	-0.09 (-0.40, 0.22)	-0.07 (-0.38, 0.25)	-
Precipitation	0.077 (-0.17, 0.32)	-0.12 (-0.14, 0.38)	0.19 (-0.06, 0.45)	0.15 (-0.10, 0.42)	0.14 (-0.11, 0.40)	0.08 (-0.17, 0.34)	-
Number Sampled	**-0.38 (-0.67, -0.10)**	**-0.37 (-0.65, -0.09)**	-0.24 (-0.49, 0.02)	-0.22 (-0.47, 0.03)	-0.24 (-0.50, 0.02)	-0.21 (-0.47, 0.05)	-
Garden	**-0.49 (-0.91, -0.08)**	**-0.50 (-0.92, -0.08)**	**-0.48 (-0.91, -0.07)**	-0.40 (-0.83, 0.002)	**-0.43 (-0.87, -0.01)**	**-0.46 (-0.90, -0.02)**	-
R2	0.037	0.036	0.025	0.023	0.024	0.023	*0.01*
kfold R2	0.022	0.0197	0.0122	0.0115	0.0108	0.0076	*0.0074*
ELPD_diff	0	-1.1	-2.6	-3	-3.9	-17.1	*-10.8*

LIST OF REFERENCES

Anandappa, J., Stanford, H., Marek, L., Goolsby, E. & Mason, C. (2023). Bioprospecting for improved floral fragrance in wild sunflowers. *Helia, 46*(79), 169-186. https://doi.org/10.1515/helia-2023-0008

Bahmani, K., Robinson, A., Majumder, S., LaVardera, A., Dowell, J. A., Goolsby, E. W., & Mason, C. M. (2022). Broad diversity in monoterpene–sesquiterpene balance across wild sunflowers: Implications of leaf and floral volatiles for biotic interactions. *American Journal of Botany*, *109*(12). https://doi.org/10.1002/ajb2.16093

Benelli, G., Canale, A., Romano, D., Flamini, G., Tavarini, S., Martini, A., Ascrizzi, R., Conte, G., Mele, M., & Angelini, L. G. (2017). Flower scent bouquet variation and bee pollinator visits in *Stevia rebaudiana* Bertoni (Asteraceae), a source of natural sweeteners. *Arthropod-Plant Interactions*, *11*(3). https://doi.org/10.1007/s11829-016-9488-y

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*. https://doi.org/10.18637/jss.v080.i01

Cariveau, D. P., Nayak, G. K., Bartomeus, I., Zientek, J., Ascher, J. S., Gibbs, J., & Winfree, R. (2016). The allometry of bee proboscis length and its uses in ecology. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0151482

Chabert, S., Mallinger, R. E., Sénéchal, C., Fougeroux, A., Geist, O., Guillemard, V., Leylavergne, S., Malard, C., Pousse, J., & Vaissière, B. E. (2022). Importance of maternal resources in pollen limitation studies with pollinator gradients: A case study with sunflower. *Agriculture, Ecosystems and Environment*, *330*. https://doi.org/10.1016/j.agee.2022.107887

Corbet, S. A., Fussell, M., Ake, R., Fraser, A., Gunson, C., Savage, A., & Smith, K. (1993). Temperature and the pollinating activity of social bees. *Ecological entomology*, *18*(1), 17-30.

Darwin, C. (1877). *The various contrivances by which orchids are fertilised by insects*. John Murray.

Dellinger, A. S. (2020). Pollination syndromes in the 21st century: where do we stand and where may we go? In *New Phytologist 228*(4). https://doi.org/10.1111/nph.16793

Endler, J. A. (1990). On the measurement and classification of colour in studies of animal colour patterns. *Biological Journal of the Linnean Society*, *41*(4). https://doi.org/10.1111/j.1095-8312.1990.tb00839.x

Ferguson, B., Mallinger, R. E., & Prasifka, J. R. (2021). Bee community composition, but not diversity, is influenced by floret size in cultivated sunflowers. *Apidologie*, *52*(6). https://doi.org/10.1007/s13592-021-00897-z

Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany*, *96*(10). https://doi.org/10.3732/ajb.0800395

Goolsby, E., Bruggeman, J., Ane, C. (2024). Rphylopars: Phylogenetic Comparative Tools for Missing Data and Within-Species Variation. R package version 0.3.10, <https://CRAN.R-project.org/package=Rphylopars>.

Govindarajan, M., Benelli, G. (2016). α-Humulene and β-elemene from Syzygium zeylanicum (Myrtaceae) essential oil: highly effective and eco-friendly larvicides against *Anopheles subpictus, Aedes albopictus*, and *Culex tritaeniorhynchus* (Diptera: Culicidae). *Parasitology Research 115*, 2771–2778. https://doi.org/10.1007/s00436-016-5025-2

Lemmon, E. W., M. O. McLinden, and D. G. Friend. (2017). NIST Chemistry WebBook, NIST Standard Reference Database. National Institute of Standards and Technology, Gaithersburg MD, USA.

Mallinger, R. E., & Prasifka, J. R. (2017). Bee visitation rates to cultivated sunflowers increase with the amount and accessibility of nectar sugars. *Journal of Applied Entomology*, *141*(7). https://doi.org/10.1111/jen.12375

Manjunath, K. R., Ray, S. S., & Vyas, D. (2016). Identification of indices for accurate estimation of anthocyanin and carotenoids in different species of flowers using hyperspectral data. *Remote Sensing Letters*, *7*(10). https://doi.org/10.1080/2150704X.2016.1210836

Marinković, R. (1992). Path-coefficient analysis of some yield components of sunflower (*Helianthus annuus* L.), I. *Euphytica*, *60*(3). https://doi.org/10.1007/BF00039399

Mason, C. M., Patel, H. S., Davis, K. E., & Donovan, L. A. (2017). Beyond pollinators: Evolution of floral architecture with environment across the wild sunflowers (*Helianthus*, Asteraceae). *Plant Ecology and Evolution*, *150*(2). https://doi.org/10.5091/plecevo.2017.1321

Merzlyak, M. N., Solovchenko, A. E., Smagin, A. I., & Gitelson, A. A. (2005). Apple flavonols during fruit adaptation to solar radiation: Spectral features and technique for non-destructive assessment. *Journal of Plant Physiology*, *162*(2). https://doi.org/10.1016/j.jplph.2004.07.002

Muchhala, N., Johnsen, S., & Smith, S. D. (2014). Competition for hummingbird pollination shapes flower color variation in andean solanaceae. *Evolution*, *68*(8). https://doi.org/10.1111/evo.12441

Oertli, S., Müller, A., & Dorn, S. (2005). Ecological and seasonal patterns in the diversity of a species-rich bee assemblage (Hymenoptera: Apoidea: Apiformes. *European Journal of Entomology*, *102*(1). https://doi.org/10.14411/eje.2005.008

Ollerton, J., Alarcon, R., Waser, N. M., Price, M. V., Watts, S., Cranmer, L., Hingston, A., Peter, C. I., & Rotenberry, J. (2009). A global test of the pollination syndrome hypothesis. *Annals of Botany*, *103*(9). https://doi.org/10.1093/aob/mcp031

Paradis. E., Schliep, K. (2019). “ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.” _Bioinformatics_, *35*, 526-528. doi:10.1093/bioinformatics/bty633

Portlas, Z. M., Tetlie, J. R., Prischmann-Voldseth, D., Hulke, B. S., & Prasifka, J. R. (2018). Variation in floret size explains differences in wild bee visitation to cultivated sunflowers. *Plant Genetic Resources: Characterisation and Utilisation*, *16*(6). https://doi.org/10.1017/S1479262118000072

Posey, A. F., Katayama, R. W., Burleigh, J. G. (1986). The Abundance and Daily Visitation Patterns of Bees (Hymenoptera: Apoidea) on Oilseed Sunflower, *Helianthus annuus* L., in Southeastern Arkansas. *Journal of the Kansas Entomological Society 59*(3)

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. In *Ecography 40*(8). https://doi.org/10.1111/ecog.02881

Rogers, C. E. (1988). Insects from native and cultivated sunflowers(*Helianthus*) in southern latitudes of the United States*. Journal of Agricultural Entomology 5*(4)

Schiestl, F. P. (2015). Ecology and evolution of floral volatile-mediated information transfer in plants. *New Phytologist*, *206*(2). https://doi.org/10.1111/nph.13243

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods 9*.

Sivula, T., Magnusson, M., & Vehtari, A. (2023). Unbiased estimator for the variance of the leave-one-out cross-validation estimator for a Bayesian normal model with fixed variance. *Communications in Statistics - Theory and Methods*, *52*(16). https://doi.org/10.1080/03610926.2021.2021240

Smith, S. D. (2014). Quantifying color variation: Improved formulas for calculating hue with segment classification. *Applications in Plant Sciences*, *2*(3). https://doi.org/10.3732/apps.1300088

Specht, C. D., & Bartlett, M. E. (2009). Flower evolution: The origin and subsequent diversification of the angiosperm flower. *Annual Review of Ecology, Evolution, and Systematics*, *40*. https://doi.org/10.1146/annurev.ecolsys.110308.120203

Sprengel. C. K. (1793) Das Entdeckte Geheimniss der Natur im Bau und in der Befruchtung der Blumen. *bei Friedrick Vieweg dem ältern*, Berlin.

Stephens, J. D., Rogers, W. L., Mason, C. M., Donovan, L. A., & Malmberg, R. L. (2015). Species tree estimation of diploid Helianthus (Asteraceae) using target enrichment . *American Journal of Botany*, *102*(6). https://doi.org/10.3732/ajb.1500031

Thompson, W. R. (1935). On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to Sample Standard Deviation. *The Annals of Mathematical Statistics*, *6*(4). https://doi.org/10.1214/aoms/1177732567

Todesco, M., Bercovich, N., Kim, A., Imerovski, I., Owens, G. L., Ruiz, Ó. D., Holalu, S. V., Madilao, L. L., Jahani, M., Légaré, J. S., Blackman, B. K., & Rieseberg, L. H. (2022). Genetic basis and dual adaptive role of floral pigmentation in sunflowers. *ELife*, *11*. https://doi.org/10.7554/eLife.72072

Vehtari, A., Gelman, A., & Gabry, J. (2017). Loo package: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models. *Statistics and Computing*, *27*(5).

WeatherSTEM. https://orange.weatherstem.com/data?refer=/data (accessed 2024-02-22).