

TIME AND TERPENOIDS: EXPERIMENTAL AND DATA-INTENSIVE INVESTIGATIONS
INTO TEMPORAL ECOLOGY AND PHYTOCHEMISTRY

By

Daniel Brendan Turner

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Entomology—Doctor of Philosophy
Ecology, Evolutionary Biology and Behavior—Dual Major

2023

ABSTRACT

Time and terpenoids mediate interactions between plants and their environment. Time has increasingly been recognized as a source of variability in ecological interactions. Despite known temporal variability in natural systems, ecological experiments often evaluate treatment effects at a single moment or as the aggregation of many moments in a community. I hypothesize that both treatment timing (i.e., when we apply a change to a system) and observation timing (i.e., when we measure a response) will affect the responses measured in herbivory experiments. Studying the interactions between *Solidago altissima* and insects and pathogens in *in situ* and common garden experiments, I evaluate how the timing and frequency of herbivory, as well as the timing of observations, affects estimations of plant growth and community responses. Feeding by *Slaterocoris* sp. (Hemiptera: Miridae) variably impacted subsequent chewing herbivory, pathogen damage, and plant height; pathogen damage was generally reduced in mirid-fed plants until the final observation date, and plant height was reduced in mirid-fed plants at all observation dates. In a second experiment, I found that late-timed jasmonic acid sprays significantly reduced chewing herbivory damage at the first observation, but not ten days later. Multiple sprays had cumulative effects on pathogen susceptibility, depending on spray timing.

Terpenoids are the most diverse group of phytochemicals, yet identifying macro-scale trends in their diversity across environmental gradients and phylogeny has lagged due to limitations in cross-study synthesis and the logistical constraints of analyzing plant tissues at a global scale. Through a meta-analysis of studies on more than 200 plant species, I tested how terpenoid diversity varies across a gradient of two climate variables—mean annual temperature and annual precipitation—and if plant species that are more closely related produce more similar terpenoid profiles. I focused on two easily detectable superclasses of terpenoids, monoterpenoids

and sesquiterpenoids. Both compound richness and structural α -diversity increased with increasing annual precipitation. I also found that more different temperature and precipitation regimes are associated with increasingly distinct terpenoid profiles. These patterns may be explained by the physicochemical properties that govern terpenoid release from plant tissue, including stomatal conductance, but further mechanistic investigations are needed. More closely related species produced more similar terpenoid profiles but less similar chemical substructures than more distantly related species. These phylogenetic patterns may be explained by plants sharing evolutionary history that constrains the overall terpenoid profile. Closely related species may also differentiate themselves through chemical substructures that govern other organisms' structurally specific perception to compounds.

Assembling large, detailed trait databases is imperative for the advancement of testing macroecological and macroevolutionary hypotheses, and most plant trait databases are assembled around morphological and life history traits that can be measured readily across many plant species. Phytochemicals have gone largely overlooked in the generation of these databases because the methods for detection, identification, and reporting vary widely across investigators and focal groups of chemicals. To lay a foundation for the development of global, quantitative, and methodologically detailed phytochemical databases, I present *terpr v1.0.0*, a database of plant monoterpenoids and sesquiterpenoids from 1127 studies containing 5107 samples from 1178 plant species, constituting 1852 unique identified monoterpenoids and sesquiterpenoids. I collected 86 features, including table indices, for each study, and present the database for further investigation into the patterns of terpenoid diversity, identification of valuable natural products, and best methodological practices.

Copyright by
DANIEL BRENDAN TURNER
2023

To the students from whom academia took their joy, may you find genuine kindness and happiness wherever your journey leads you.

ACKNOWLEDGEMENTS

I begin by acknowledging the person without whom I would have abandoned the dissertation long ago—Jordi Rivera Prince. She is my sounding board, my support, and my happiness, especially when we were stuck alone inside for many months. Without her, I would have a third of the data to show in my first chapter, and I would still be straining my eyes at the computer screen for the second and third. I am so grateful for you and Rocoto and Ceviche, more than I have words to describe.

To my sister, Elizeth Cinto Mejía, you and I have overcome so much together, and I can never repay the years of support and laughs we shared throughout the years, except maybe through a pair of Manolos after we make it big. To Kayleigh Hauri, thank you for keeping me grounded since day one on 39th Street and always sticking up for what is just. To Andrea Glassmire, Joshua Snook, and Rusty, thank you for keeping our family together through the long, dark winters. In addition to being the most generous and strongest of humans, these people are excellent scientists and inspired much of my work.

Another group of people rallied around me through their friendship and support, whether visiting me from thousands of miles away or sharing a drink with me a couple blocks nearby. Thank you, Gaelen McCartney, Daron Mulligan, Andrew Drake, Dennis Parnell, Isis Dwyer, Morgan Cassidy, Stephanie David, Amanda Brock, David Drennan, Aldo Watanave, and Shelly Watanave for your friendship. Thanks to my friends and colleagues from MSU, especially Luke Zehr, Rachel Osborn, and Natalie Constancio.

Thank you to my dissertation chair, Rufus Isaacs, for helping me across the finish line. My conversations with Diego Salazar were vital to my terpenoid work and the ideas presented here. Thank you to Lars Brudvig and Marianna Szucs for serving on this committee since the

early days of this degree. Thank you to Keri Greig, Carolyn Graham, and the rest of the terpenoid data group, for your help collecting data and important contributions to the project's structure. Bryan Jurado, Leah Flores Cabrera, Minali Bhatt, Kelsey Doud, and Brendan Randell were amazing undergraduates to work with, and they taught me as much as or more than I taught them. To Heather Lenartson-Kluge, Linda Gallagher, and the entire administrative staff in the Department of Entomology, you all do more than a student like me sees firsthand, and I am so grateful for you. Thanks to the MSU Graduate Employees Union for negotiating my health insurance. The terpenoid project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2018-67013-28065 from the USDA National Institute of Food and Agriculture.

Thank you to my mentors from (geographically) afar, Jocelyn Behm and Matthew Helmus, for guiding me through my first first-author publication in such a meaningful framework and for sparking my initial interest in ecological data. Thank you to Adela Ruiz, who gave me the opportunity to put my statistics knowledge to practical use, taught me how to report complicated numbers powerfully, and shared so much joy and laughter.

Thank you to the plants and insects, who are more than just the numbers we value so much in Western ecology.

Finally, thank you to my parents who financially and personally sacrificed so much for my formal education. My mom and I attended graduate school at the same time, and I still swell with pride thinking about her earning that Masters in Public History in 2021. My dad, always the hands-on carpenter, was only a phone call away when I was entirely in-over-my-head constructing something in the field. Thanks to Maryellen Blake for always believing in me and sparking a love for science at Rocket Day. To my family in Michigan, Jessica, Pieter, Beth,

Fabio, Sydney, Josh, Diego, Emilio, and Carlo, it was always a comfort knowing you were only a short drive away. To the people who passed down their love for plants, “the oldest members of our community,” without even knowing, thank you.

To all the people who made this process less lonely and more joyful, I cannot thank you enough.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
REFERENCES	9
CHAPTER 2: TEMPORAL CONTEXT OF HERBIVORY AFFECTS GOLDENROD COMMUNITY ECOLOGY AND PLANT GROWTH	13
REFERENCES	31
CHAPTER 3: PLANT TERPENOID DIVERSITY VARIES WITH TEMPERATURE, PRECIPITATION, AND PHYLOGENY: A META-ANALYSIS	40
REFERENCES	63
CHAPTER 4: TERPR V1.0.0: A DATABASE OF PLANT MONOTERPENOID AND SESQUITERPENOID	69
REFERENCES	94
CHAPTER 5: CONCLUSION	98
REFERENCES	106
APPENDIX 1: RECORD OF DEPOSITION OF VOUCHER SPECIMENS	107
APPENDIX 2: SUPPLEMENT FOR TEMPORAL CONTEXT OF HERBIVORY AFFECTS GOLDENROD COMMUNITY ECOLOGY AND PLANT GROWTH	108

CHAPTER 1:

INTRODUCTION

The ecology and diversity of plant-insect interactions are shaped by processes that humans cannot immediately observe. These processes are omnipresent in the natural world, from the sequestration of toxic plant cardiac glycosides in the gut of the milkweed bug, *Oncopeltus fasciatus* (Scudder and Meredith 1982) to the chewing of galling insects inside plant stems (Hartnett and Abrahamson 1979, Fay *et al.* 1996, Hammer *et al.* 2021) to parasitoids' reception of chemical cries for help from plants (Turlings *et al.* 1991, Kroes *et al.* 2017, Da Silva *et al.* 2022). We know they occur in nature, but we often observe the results of these processes, not the processes *per se*. We witness the aposematic coloration of *O. fasciatus* indicating they are unpalatable to predators, the tumorous growth of a gall after the plant reacts to its internal infestation, or the zombification of a caterpillar filled with parasitoid eggs after a leafy meal. However, many factors—the timing, place, and specific chemical reactions of the ecological interaction—must coalesce before we even observe these phenomena. By challenging and expanding existing experimental, meta-analytical, synthetic approaches in ecology, we can begin to understand how organisms respond to stimuli in various timescales and across different spatial extents to reveal the chemical underpinnings of biological and ecological diversity.

In this dissertation, I focus on two factors—time and chemistry—and how they influence ecological and biological variation, specifically in plants and their community interactions with insects. Time is a fundamental axis of ecological interactions. Given shifts in organisms' phenologies with climate change, ecologists are challenged with understanding how organisms' internal clocks and calendars are changing from historical norms (Piao *et al.* 2019, Lian *et al.* 2020). From recent analyses, it is increasingly clear that the timing-based effects of climate

change cascade from individuals to populations, communities, and ecosystems (Cinto Mejía and Wetzel 2023). For instance, flowers now bloom asynchronously with the timing of pollinator activity (Kudo and Ida 2013), and insect pests undergo more rapid population growth with increased generations per year (Rao *et al.* 2015). Thus, we need deeper investigations into the temporal dynamics of basic and applied ecological systems to understand how and how quickly our world is changing and the implications for the ecology of our natural world.

Changes in systems, such as biotic and abiotic stressors, do not occur at a single time point, but rather across many time points that accumulate within and across growing seasons (Ryo *et al.* 2019, Jackson *et al.* 2021). Therefore, capturing an annual mean, maximum, minimum, or final value along the period of study is insufficient for quantifying and describing ecological processes. Capturing the full arc of an organism's growth and interactions with other organisms will require modeling repeated measurements of the same system. Long-term ecological experiments have been doing this for many years (Cusser *et al.* 2021, Resasco *et al.* 2023). Yet, in every ecological question or hypothesis, time is an implicit, if not explicit, force behind observed patterns. Some conditions that change through time are more easily quantifiable, including precipitation and temperature, tissue maturation, and nutrient availability. Other conditions are more difficult to measure, including unique life history events and intraspecific and interspecific encounters.

Regardless of quantifiability, these events can alter the trajectories of organisms, populations, and communities through changes in biochemistry and resource competition. For example, insect chewing herbivory, a seemingly isolated one-time event, represents a life history event whose influence can cascade across trophic levels beyond those of the focal plant and insect (Dicke and van Loon 2000, Orre Gordon *et al.* 2013). Whether the herbivore chose its host

at random or not, there are some irreparable changes in the community—a parasitoid may perceive chemical cues from the plant to find its herbivore host, the plant may send signals to its plant neighbors to increase defenses, or the plant itself may have induced chemical responses, such as stimulating the production of jasmonic acid and its derivative methyl jasmonate (Baldwin 1998; Sirhindi *et al.* 2020), altering investment in defense and growth. Still, these individual and community responses do not occur at a single time point, and future experiments will require repeated measurements to determine trajectories across scales.

From flavonoids to fatty acids, chemistry rules the world of plants and insects. By the number of identified unique compounds, terpenoids are the most diverse group of plant chemicals (Gershenzon and Dudareva 2007). They are also known as isoprenoids and are united by their five-carbon building blocks and series of chemical reactions and synthases, as described in Chapter 3 (Chen *et al.* 2011). Many terpenoids are the byproducts of plant primary metabolism and so they are referred to as secondary metabolites. Despite this terminology, these terpenoid secondary metabolites are hardly the waste products they were once considered. Isoprene, the simplest isoprenoid, is responsible for nearly 50% of all biogenic volatile organic compound emissions from plants (Guenther *et al.* 2012), significantly contributes to Earth's overall atmospheric chemistry (Sharkey *et al.* 2008), and can protect plant tissues from heat stress along with its ten-carbon chain counterparts, monoterpenoids (Sharkey *et al.* 2001; Copolovici *et al.* 2005).

Lighter weight terpenoids, such as monoterpenoids and 15-carbon sesquiterpenoids are found across foliar, floral, and root, among other tissues. They can be released in the gaseous phases at the leaf surface or through the stomata allowing for future reception by other plants, insects, fungi, and microbes (Niinemets *et al.* 2004). They can also be ingested by insects,

mammals, and birds in aqueous phase, thus constituting a vital source of flavor (Schwab *et al.* 2008). Monoterpenoids and sesquiterpenoids are also part of a broader non-volatile and volatile bouquet of secondary metabolites that plants produce to defend themselves against antagonism, attract mutualists, and communicate with other plants (Kessler and Kalske 2018).

The abiotic and biotic conditions promoting the diversity of monoterpenoids and sesquiterpenoids have been the subject of many hypotheses in plant-insect ecology and evolution, with greater emphasis toward the biotic interactions that promote selection for one compound, or one configuration of a compound, over another. One of the enduring hypotheses, the escape-and-radiate hypothesis (Ehrlich and Raven 1964) posits that plants evolve novel defenses to protect against insect antagonists, and in turn, insect antagonists evolve novel mechanisms to sequester, evade, or otherwise neutralize these defenses. We may observe more closely related species to have more similar chemical profiles (i.e., display phylogenetic signal), such as phenylpropanoid glycosides in *Mimulus* leaves (Holeski *et al.* 2020), and differentiate themselves from near relatives by only a few compounds. Alternatively, defenses may evolve independently of phylogeny, as in the case of *Inga* saplings from the Peruvian Amazon (Endara *et al.* 2017). Many other hypotheses focus on the effects of phytochemical diversity *per se*. For example, the interaction-diversity hypothesis suggests that plants require a multitude of chemicals because they interact with many other species (Iason *et al.* 2011). A recent experiment found support for this interaction-diversity hypothesis when examining lepidopteran responses to phenolic metabolites (Whitehead *et al.* 2021).

Nevertheless, macro-scale analyses testing these hypotheses have lagged behind system-specific studies, especially those that also evaluate the abiotic conditions constraining or promoting phytochemical diversity. Formal incorporation of the environmental conditions that

constrain or give rise to phytochemical diversity is a promising avenue of research that has not been visited in earnest since the mid-2000s. At that time, specific isoprenoids, including isoprene, monoterpenoids, and sesquiterpenoids were found to vary with annual mean temperature (Llusà *et al.* 2006), but recent research has found phytochemical diversity to correlate with climatic, geographical, and phylogenetic covariates across many phytochemical groups (Defosse *et al.* 2021).

Several physiological and physicochemical mechanisms may explain macro-scale trends in phytochemical diversity. For example, stomatal conductance correlates with monoterpenoid emissions based on the difference between the internal plant tissue partial pressure and the atmosphere (Niinemets *et al.* 2002). When oxygenated monoterpenoids are accumulated in the plant tissue, changes in this pressure differential upon stomatal opening promote the emissions of these hydrophilic compounds that would otherwise continue to accumulate internally (Niinemets *et al.* 2004). The temperature-dependence of monoterpenoid and sesquiterpenoid emissions from plants is another example (Helmig *et al.* 2013). Despite the many mechanisms that could explain phytochemical diversity, especially for monoterpenoids and sesquiterpenoids, hypotheses and analyses have not fully embraced incorporating these abiotic covariates and so we have limited understanding of the large-scale patterns of how plant secondary metabolites vary with environmental conditions.

Identifying broad diversity patterns in any group of phytochemicals is limited by several logistical hurdles. Collecting and analyzing plant tissue with enough geographic and phylogenetic coverage is an obvious obstacle due to the logistics of this scale of field work. Synthetic approaches that collate information across publications has historically been ruled out due to variation in compound identification methods, the different compound naming

conventions in different disciplines, and the sheer volume of unstructured (i.e., not tabular form that is ready for analysis) data provided by these studies. However, newer software, cloud computing capacity, and programming tools have untapped potential to bring these data together for the first time so they can be analyzed, existing hypotheses can be refined and tested, and new hypotheses can be generated.

The proliferation of functional trait databases has shifted priorities, general theories, and research accessibility in ecology, and existing plant trait databases can serve as models for phytochemical databases. Many of these databases have assembled user-submitted data that is structured and standardized by the database authors, such as *FunAndes* (Báez *et al.* 2022) and the *TRY* plant trait database (Kattge *et al.* 2011). As of 2019, data from *TRY* has contributed to over 250 publications (Kattge *et al.* 2020) testing hypotheses ranging from the relationship between life form and allelopathy (Zhang *et al.* 2021) to the effects of anthropogenic stress on plant height (Newbold *et al.* 2015). In the latest major publication about *TRY*, the authors discussed the incorporation of legacy and published trait data into the future iterations of the database (Kattge *et al.* 2020). The challenges to assembling a phytochemical database in this way are also faced in the development of general plant trait databases—variation in measurement methodology and reporting, bias in studies toward certain geographical regions or species groups, and resource limitations to process unstructured data. However, database engineering and deployment protocols, such as findable, accessible, interoperable, and reusable (FAIR) (Jacobsen *et al.* 2020) and Open Science (Gallagher *et al.* 2020) principles, provide useful guidance for the next generation of databases.

With the expansion of these principles and technical tools that harness the power of large datasets, we can now develop techniques to study global patterns in phytochemical diversity that

standardize the many methodologies through which compounds are identified and reported. Plant monoterpenoids and sesquiterpenoids are well suited to beginning this endeavor. Extracted through the distillation of plant tissue or adsorbed from the headspace of the plant, monoterpenoids and sesquiterpenoids are two superclasses (i.e., subcategories of compounds that share biosynthetic pathways) of secondary plant metabolites that are routinely identified using coupled gas chromatography and mass spectrometry (GC-MS). In this method, compounds are vaporized into the GC, carried using a gas through a capillary column, and identified by the retention time or index relative to an internal standard that is based on their physicochemical properties (Sparkman *et al.* 2011). In a GC-MS system the compounds are also ionized in the MS system resulting in more information about the structure and identity of the compound and its molecular structure after referencing existing MS libraries (Sparkman *et al.* 2011). Monoterpenoids and sesquiterpenoids have a smaller molecular weight and are more volatile than many other groups of phytochemicals, presenting an ideal model for the collection, curation, and analysis of phytochemistry across many published studies.

Explicitly evaluating variation in species interactions throughout time and synthesizing the scientific community's existing data on terpenoids has the potential to reveal overlooked processes and factors underlying ecological diversity. In this dissertation, I coordinated a variety of approaches to achieve these goals, from experimentation to data synthesis and analysis. First, I deployed a common garden experiment and an *in situ* field experiment with a perennial plant, *Solidago altissima*, and its antagonistic insect and pathogen community. In these experiments, I tested how the timing of experimental treatments and response variable measurements influences observed variability in ecological interactions. Second, I address the climatic and phylogenetic relationships with chemical diversity in plant monoterpenoids and sesquiterpenoids through a

meta-analysis from studies of 206 plant species conducted around the globe. Finally, I assemble and present the largest, most high resolution database of monoterpenoids and sesquiterpenoids, to facilitate future analysis into the causes and effects of plant secondary metabolite diversity.

REFERENCES

- Báez S, Cayuela L, Macía MJ, et al (2022) FunAndes – A functional trait database of Andean plants. *Sci Data* 9:511. <https://doi.org/10.1038/s41597-022-01626-6>
- Baldwin IT (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proc Natl Acad Sci USA* 95:8113–8118. <https://doi.org/10.1073/pnas.95.14.8113>
- Beran F, Petschenka G (2022) Sequestration of plant defense compounds by insects: From mechanisms to insect–plant coevolution. *Annual Review of Entomology* 67:163–180. <https://doi.org/10.1146/annurev-ento-062821-062319>
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* 66:212–229. <https://doi.org/10.1111/j.1365-3113.2011.04520.x>
- Cinto Mejía E, Wetzel WC (2023) The ecological consequences of the timing of extreme climate events. *Ecology and Evolution* 13:e9661. <https://doi.org/10.1002/ece3.9661>
- Copolovici LO, Filella I, Llusà J, et al (2005) The capacity for thermal protection of photosynthetic electron transport varies for different monoterpenes in *Quercus ilex*. *Plant Physiology* 139:485–496. <https://doi.org/10.1104/pp.105.065995>
- Cusser S, Helms IV J, Bahlai CA, Haddad NM (2021) How long do population level field experiments need to be? Utilising data from the 40-year-old LTER network. *Ecology Letters* 24:1103–1111. <https://doi.org/10.1111/ele.13710>
- Da Silva ITFA, Magalhães DM, Borges M, et al (2022) Exploitation of herbivore-induced cotton volatiles by the parasitic wasp *Bracon vulgaris* reveals a dominant chemotactic effect of terpenoids. *BioControl* 67:135–148. <https://doi.org/10.1007/s10526-022-10135-9>
- Defosse E, Pitteloud C, Descombes P, et al (2021) Spatial and evolutionary predictability of phytochemical diversity. *Proc Natl Acad Sci USA* 118:e2013344118. <https://doi.org/10.1073/pnas.2013344118>
- Dicke M, van Loon JJA (2000) Multitrophic effects of herbivore-induced plant volatiles in an evolutionary context. *Entomologia Experimentalis et Applicata* 97:237–249. <https://doi.org/10.1046/j.1570-7458.2000.00736.x>
- Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18:586–608. <https://doi.org/10.1111/j.1558-5646.1964.tb01674.x>
- Endara M-J, Coley PD, Ghabash G, et al (2017) Coevolutionary arms race versus host defense chase in a tropical herbivore–plant system. *Proc Natl Acad Sci USA* 114:. <https://doi.org/10.1073/pnas.1707727114>

- Fay PA, Hartnett DC, Knapp AK (1996) Plant tolerance of gall-insect attack and gall-insect performance. *Ecology* 77:521–534. <https://doi.org/10.2307/2265627>
- Gallagher RV, Falster DS, Maitner BS, et al (2020) Open Science principles for accelerating trait-based science across the tree of life. *Nat Ecol Evol* 4:294–303. <https://doi.org/10.1038/s41559-020-1109-6>
- Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology* 3:408–414. <https://doi.org/10.1038/nchembio.2007.5>
- Guenther AB, Jiang X, Heald CL, et al (2012) The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development* 5:1471–1492. <https://doi.org/10.5194/gmd-5-1471-2012>
- Hammer TJ, De Clerck-Floate R, Tooker JF, et al (2021) Are bacterial symbionts associated with gall induction in insects? *Arthropod-Plant Interactions* 15:1–12. <https://doi.org/10.1007/s11829-020-09800-6>
- Hartnett DC, Abrahamson WG (1979) The effects of stem gall insects on life history patterns in *Solidago canadensis*. *Ecology* 60:910–917. <https://doi.org/10.2307/1936859>
- Helmig D, Daly RW, Milford J, Guenther A (2013) Seasonal trends of biogenic terpene emissions. *Chemosphere* 93:35–46. <https://doi.org/10.1016/j.chemosphere.2013.04.058>
- Holeski LM, Keefover-Ring K, Sobel JM, Kooyers NJ (2021) Evolutionary history and ecology shape the diversity and abundance of phytochemical arsenals across monkeyflowers. *Journal of Evolutionary Biology* 34:571–583. <https://doi.org/10.1111/jeb.13760>
- Iason GR, O'Reilly-Wapstra JM, Brewer MJ, et al (2011) Do multiple herbivores maintain chemical diversity of Scots pine monoterpenes? *Philos Trans R Soc Lond B Biol Sci* 366:1337–1345. <https://doi.org/10.1098/rstb.2010.0236>
- Jackson MC, Pawar S, Woodward G (2021) The temporal dynamics of multiple stressor effects: From individuals to ecosystems. *Trends in Ecology & Evolution* 36:402–410. <https://doi.org/10.1016/j.tree.2021.01.005>
- Jacobsen A, de Miranda Azevedo R, Juty N, et al (2020) FAIR Principles: Interpretations and implementation considerations. *Data Intelligence* 2:10–29. https://doi.org/10.1162/dint_r_00024
- Kattge J, Bönisch G, Díaz S, et al (2020) TRY plant trait database – enhanced coverage and open access. *Global Change Biology* 26:119–188. <https://doi.org/10.1111/gcb.14904>
- Kattge J, Díaz S, Lavorel S, et al (2011) TRY – a global database of plant traits. *Global Change Biology* 17:2905–2935. <https://doi.org/10.1111/j.1365-2486.2011.02451.x>

- Kessler A, Kalske A (2018) Plant secondary metabolite diversity and species interactions. *Annu Rev Ecol Evol Syst* 49:115–138. <https://doi.org/10.1146/annurev-ecolsys-110617-062406>
- Kroes A, Weldegergis BT, Cappai F, et al (2017) Terpenoid biosynthesis in *Arabidopsis* attacked by caterpillars and aphids: effects of aphid density on the attraction of a caterpillar parasitoid. *Oecologia* 185:699–712. <https://doi.org/10.1007/s00442-017-3985-2>
- Kudo G, Ida TY (2013) Early onset of spring increases the phenological mismatch between plants and pollinators. *Ecology* 94:2311–2320. <https://doi.org/10.1890/12-2003.1>
- Lian X, Piao S, Li LZ, et al (2020) Summer soil drying exacerbated by earlier spring greening of northern vegetation. *Science Advances* 6:eaax0255. <https://doi.org/10.1126/sciadv.aax0255>
- Llusà J, Peñuelas J, Sardans J, et al (2010) Measurement of volatile terpene emissions in 70 dominant vascular plant species in Hawaii: aliens emit more than natives. *Global Ecology and Biogeography* 19:863–874. <https://doi.org/10.1111/j.1466-8238.2010.00557.x>
- Newbold T, Hudson LN, Hill SLL, et al (2015) Global effects of land use on local terrestrial biodiversity. *Nature* 520:45–50. <https://doi.org/10.1038/nature14324>
- Niinemets U (2002) Stomatal constraints may affect emission of oxygenated monoterpenoids from the foliage of *Pinus pinea*. *Plant Physiology* 130:1371–1385. <https://doi.org/10.1104/pp.009670>
- Niinemets Ü, Loreto F, Reichstein M (2004) Physiological and physicochemical controls on foliar volatile organic compound emissions. *Trends in Plant Science* 9:180–186. <https://doi.org/10.1016/j.tplants.2004.02.006>
- Orre Gordon GUS, Wratten SD, Jonsson M, et al (2013) ‘Attract and reward’: Combining a herbivore-induced plant volatile with floral resource supplementation – Multi-trophic level effects. *Biological Control* 64:106–115. <https://doi.org/10.1016/j.biocontrol.2012.10.003>
- Piao S, Liu Q, Chen A, et al (2019) Plant phenology and global climate change: Current progresses and challenges. *Global Change Biology* 25:1922–1940. <https://doi.org/10.1111/gcb.14619>
- Resasco J, Burt MA, Orrock JL, et al (2023) Transient effects of corridors on polygyne fire ants over a decade. *Ecological Entomology* 48:263–268. <https://doi.org/10.1111/een.13214>
- Ryo M, Aguilar-Trigueros CA, Pinek L, et al (2019) Basic principles of temporal dynamics. *Trends in Ecology & Evolution* 34:723–733. <https://doi.org/10.1016/j.tree.2019.03.007>
- Schwab W, Davidovich-Rikanati R, Lewinsohn E (2008) Biosynthesis of plant-derived flavor compounds. *The Plant Journal* 54:712–732. <https://doi.org/10.1111/j.1365-313X.2008.03446.x>

- Scudder GGE, Meredith J (1982) The permeability of the midgut of three insects to cardiac glycosides. *Journal of Insect Physiology* 28:689–694. [https://doi.org/10.1016/0022-1910\(82\)90147-0](https://doi.org/10.1016/0022-1910(82)90147-0)
- Sharkey TD, Chen X, Yeh S (2001) Isoprene increases thermotolerance of fosmidomycin-fed Leaves. *Plant Physiology* 125:2001–2006. <https://doi.org/10.1104/pp.125.4.2001>
- Sharkey TD, Wiberley AE, Donohue AR (2008) Isoprene emission from plants: Why and how. *Annals of Botany* 101:5–18. <https://doi.org/10.1093/aob/mcm240>
- Sirhindi G, Mushtaq R, Gill SS, et al (2020) Jasmonic acid and methyl jasmonate modulate growth, photosynthetic activity and expression of photosystem II subunit genes in *Brassica oleracea* L. *Sci Rep* 10:9322. <https://doi.org/10.1038/s41598-020-65309-1>
- Sparkman OD, Penton Z, Fulton K (2011) Gas chromatography and mass spectrometry: A practical guide, 2nd edn. Academic Press, Burlington, MA
- Srinivasa Rao M, Swathi P, Rama Rao CA, et al (2015) Model and scenario variations in predicted number of generations of *Spodoptera litura* Fab. on peanut during future climate change scenario. *PLoS ONE* 10:e0116762. <https://doi.org/10.1371/journal.pone.0116762>
- Turlings TCJ, Tumlinson JH, Heath RR, et al (1991) Isolation and identification of allelochemicals that attract the larval parasitoid, *Cotesia marginiventris* (Cresson), to the microhabitat of one of its hosts. *J Chem Ecol* 17:2235–2251. <https://doi.org/10.1007/BF00988004>
- Watts K, Whytock RC, Park KJ, et al (2020) Ecological time lags and the journey towards conservation success. *Nat Ecol Evol* 4:304–311. <https://doi.org/10.1038/s41559-019-1087-8>
- Wauchope HS, Amano T, Geldmann J, et al (2021) Evaluating impact using time-series data. *Trends in Ecology & Evolution* 36:196–205. <https://doi.org/10.1016/j.tree.2020.11.001>
- Whitehead SR, Bass E, Corrigan A, et al (2021) Interaction diversity explains the maintenance of phytochemical diversity. *Ecology Letters* ele.13736. <https://doi.org/10.1111/ele.13736>
- Wolkovich EM, Cook BI, McLauchlan KK, Davies TJ (2014) Temporal ecology in the Anthropocene. *Ecology Letters* 17:1365–1379. <https://doi.org/10.1111/ele.12353>
- Zhang Z, Liu Y, Yuan L, et al (2021) Effect of allelopathy on plant performance: a meta-analysis. *Ecology Letters* 24:348–362. <https://doi.org/10.1111/ele.13627>

CHAPTER 2:
TEMPORAL CONTEXT OF HERBIVORY AFFECTS GOLDENROD COMMUNITY
ECOLOGY AND PLANT GROWTH

Abstract

Organisms can undergo vast physiological and ecological changes at short time scales, yet few studies in ecology examine the temporal pattern of organismal responses to interspecific encounters. By varying treatment timings and repeatedly sampling the same individuals and communities within a season, we can track the trajectory of species interactions and changes in development. Using the tall goldenrod (*Solidago altissima*)-herbivore-pathogen community, we varied two herbivory treatments, foliar mirid feeding and exogenous jasmonic acid (JA) sprays, in time and repeatedly measured indicators of plant growth and species interactions across the *S. altissima* ontogeny. The timing and frequency of herbivory treatments resulted in significant changes in antagonist damage and plant growth, but effects often varied by observation date. After mirid-feeding, we initially found reduced pathogen damage followed by 38% increase 66 days later. Relative to control plants, mirid-fed plant height was 11% lower at the first observation date but only 1% lower by the last. In the JA spray experiment, we initially found a 34% reduction in chewing herbivory damage in plants that were sprayed with JA at a late timing relative to control plants but no significant difference ten days later. We observed a 97% increase in chewing herbivory in plants that were sprayed with JA both early and in the middle of the experiment relative to plants that were only sprayed in the middle timing. Our findings demonstrate the effects of herbivory on community dynamics and plant-growth depend on timing and frequency.

Introduction

Time is a fundamental axis on which ecologists explore the patterns of the natural world. Population dynamics, phenology, and ontogeny require an explicit use of time to describe ecological phenomena. Important ecological events in an organism's life history include the sudden presence of a competitor, predator attack, and both gradual and rapid changes in environmental conditions. Some events can cause priority effects—when a species' early arrival in a community alters community development, for example, by shifting resource quality or availability for later arriving species (Fukami 2015). These ecological events can have varying degrees of impact based on their timing relative to the scale of the organism's life cycle and the trajectory of the ecological community it interacts with (Wauchope et al. 2021; Jackson et al. 2021).

An organism's response to environmental change can dampen, strengthen, or remain static as time progresses, but the effects in experimental studies are often measured at a single moment or aggregated across time points (Yang 2020). This can hide temporal dynamics that are important for understanding interactions within the community and the abiotic environment. In experimental ecology, timing can be split into two factors: treatment timing and observation timing. Treatment timing can be described as when some organism undergoes a controlled change in their biotic or abiotic environment, while observation timing is when the human observer measures some response in the organism or its environment. Studies applying treatments at different points along organisms' ontogeny or a community ecological trajectory found variable results across different treatment timings (Barton 2013; Yang et al. 2020; Rasmussen and Yang 2022), and some even account for variation in both treatment and observation timing (Clark and Johnston 2011; Wang et

al. 2018; van Dijk et al. 2020). Without accounting for observational and treatment timings in experimental systems, we are unable to generalize the significance of ecological encounters.

The application of repeated treatments and measurements is a solution that allows ecologists to characterize temporal dynamics. This approach is not new in experimental ecology, and important insights have been gained from long-term experiments that temporally contextualize ecological populations and communities by repeating the same treatments and/or measurements (Borer et al. 2014; Bahlai et al. 2021). Yet, the events that constitute these long-term trends are scale-dependent, such that low observation frequency can obfuscate the underlying processes behind an observed ecological phenomenon (Ryo et al. 2019). By narrowing the window of observations in time and manipulating the timing and frequency of experimental treatments, we can investigate how the granularity of species interactions scales up from shorter to longer trajectories in community ecology.

Plant-insect herbivore systems are uniquely situated for these temporal investigations. First, plant resistance or susceptibility to antagonism is often explained by ontogenetically variable investments in physical and chemical defenses, optimal growth-defense-reproduction strategies, and resource optima that provide “windows of opportunities” for interacting species (Ochoa-López et al. 2015; Rusman et al. 2020; Yang and Censer 2020). As plants partition resources and defenses in time, herbivorous insects vary in phenology for competitive, reproductive, and survival purposes (Cronin 2007; Blitzer and Welter 2011; Ekholm et al. 2020). Fortunately, many events in plant-insect ecology occur on timescales that can be manipulated and observed at multiple measurement points. For example, insect feeding, or the threat of insect feeding, can induce phytochemical changes in plants that affect subsequent species interactions (Agrawal 1998; Karban 2020). These

changes can include reduced nutrient quality and increased concentrations of secondary metabolite defenses (Nykänen and Koricheva 2004; Kant et al. 2015).

We conducted two field experiments using a plant-pathogen-herbivore system to explore the role of temporal context in plant-insect interactions by varying when we apply a treatment and when we observe a pattern. In the first experiment, we applied an herbivory treatment from a naturally occurring herbivore that arrives early in the phenology of *S. altissima*. We repeatedly sampled response variables to answer the question: *How does a single early-season herbivory event alter the temporal pattern of community interactions and plant growth?* In the second experiment, we applied jasmonic acid (JA) sprays to mimic herbivory at various timings and frequencies across the *S. altissima* growing season to answer the following two questions: *How does variation in the timing of herbivory affect the temporal pattern of community interactions and plant growth? How does variation in the frequency of herbivory affect the temporal pattern of community interactions and plant growth?*

Materials and Methods

Study system

The common flowering perennial *Solidago altissima* L. (tall goldenrod; Asteraceae) have provided a model system to understand ecological phenomena. Native to North America but naturalized or invasive globally, *S. altissima* grows in old field and early successional environments. Its herbivore community feeds through varying modes, from the specialist goldenrod gallfly, *Eurosta solidaginis* (Diptera: Tephritidae), that oviposits in mid-spring (Anderson et al. 1989) and whose larvae chew on internal stem tissue to specialist mesophyll sucking mirids (e.g., *Slaterocoris* sp.) to generalist foliar chewing insects (e.g., grasshoppers) active later in the season (Burghardt 2016). *Solidago altissima* has several systems for defense,

including a “ducking” growth morphology exhibited in some genotypes that is correlated with reduced oviposition from *E. solidaginis* (Wise and Abrahamson 2008) and herbivore-induced volatile signals that warn conspecifics of herbivore attack and thus induce resistance (Kalske et al. 2019). Importantly, *S. altissima* grows throughout the spring and summer before flowering in early fall, with several months of vegetative growth susceptible to herbivory. It reproduces both sexually by seed and asexually through rhizomes that form an underground web of connections between genetically identical aboveground shoots.

Experiment #1: Applying early-season herbivory with specialist mirid

In late Apr- and early May-2019, we collected *S. altissima* rhizomes from 17 putative genets across five populations at the Lux Arbor Forest Reserve (Prairieville Township, Michigan, USA; 42.484106, -85.451471). Genets within the same population were sampled at least fifteen meters apart to ensure genetically identical individuals were not collected multiple times inadvertently (Cain 1990, p. 198; Burghardt 2016). Rhizomes were cleaned and cut into 1.5 mL uniform pieces, including any lateral roots, as measured through volume displacement in a graduated cylinder (Pang et al. 2011). Then, we placed the rhizome in soil (Suremix, Michigan Grower Products, Inc., Michigan, USA) under greenhouse conditions. See Table S2.1 for exact coordinates, dates rhizomes were collected, and dates planted in the greenhouse. We collected a common vegetative tissue-feeding insect, *Slaterocoris* sp. (Hemiptera: Miridae), from *S. altissima* plants in late Jun-2019, in the Kellogg Experimental Forest Reserve (Augusta, Michigan, USA; 42.363827, -85.352920). We placed the mirids on 58 haphazardly selected caged plants on 03-Jul-2019, while still under greenhouse conditions, and left them to feed until 20% of all expanded leaf tissue was damaged, which lasted several hours to about a day for each plant having three to five mirids. Eighty-four plants did not have damage (controls). We measured damage through a visual

survey of the leaf surface area covered in the fine white stippling *Slaterocoris* sp. creates after feeding by sucking on the mesophyll. We reused some insects across multiple plants due to collection limitations. Voucher specimens for *Slaterocoris* sp. individuals used in this experiment are deposited in the A.J. Cook Arthropod Research Collection at Michigan State University under Voucher No. 2023-03.

Immediately after the herbivory treatments were completed on 09-Jul-2019, we transplanted unexposed control plants and mirid-damaged plants to a common garden in a former agricultural field located approximately 10.5 km away from the plants' original populations. In the common garden, the plants were transplanted into buried plastic pots (3.79 l) with the bottom removed. The pots prevented the rhizome from spreading across the common garden, and the pot bottoms were open to allow lateral roots from becoming bound. The immediate area (approximately 0.25 m radius) around each plant was weeded regularly. Each plant was 1.5 meters apart and arranged in an alternating pattern between control and mirid-damaged plants (see Figure S2.1 for a map and Table S2.2 for sample sizes at each time point).

Eight days after transplanting these plants, we conducted the first survey by collecting the following information: percent leaf tissue missing due to chewing herbivory measured through visual surveying, percent leaf surface area with pathogen damage measured through visual surveying, height from the ground to the tallest part of the plant, and the number of leaves. Visual surveys were based on all leaves of the plant, and we only measured chewing herbivory damage from arthropods. Leaves were rarely completely chewed to the petiole, but for those that were, the amount of tissue missing was estimated from nearby leaves of similar age. The pathogens present on *S. altissima* primarily consisted of leaf rust and fungal and bacterial leaf spots, and we could not account for pathogen damage that was removed due to leaf herbivory. To observe variations

in time and in response to this mirid feeding, we conducted these measurements at four time points through the season: 17-19-Jul-2019, 12-14-Aug-2019, 31-Aug-02-Sep-2019, and 21-22-Sep-2019 (Figure 2.1a). Plants were left in the common garden until we collected them on 13-Dec-2019. The rhizomes were separated from aboveground tissue, and we measured their volumes.

Experiment 2: Manipulating the timing and frequency of herbivory through jasmonic acid sprays

This experiment was conducted in a natural population of *S. altissima* at the Lux Arbor Forest Reserve, where we collected rhizomes for the mirid feeding experiment described above. The population was split into four replicate plots to ensure sufficient genetic diversity in the focal plants (See Figure 2.2 for plots on map). The centers of the plots were at least 15 m apart. Within each plot, between 29-32 plants were selected, at least 50 cm apart, and randomly assigned to one of the JA spray timings.

While several chemicals can be sprayed on plants to mimic herbivory, JA is an important regulator for plant resistance and defense to antagonism (reviewed in Howe and Jander 2008). The exogenous application of JA on plants has been shown to reduce herbivore growth rates and affect photosynthetic activity at various time scales (Thaler et al. 1996; Babst et al. 2005). Increased levels of JA can be induced in *S. altissima* through generalist herbivory and simple exposure to certain insect herbivores (Tooker and De Moraes 2009; Helms et al. 2013). At concentrations of 1.0 mM, exogenous JA application can shift growth rate and reproduction in *S. altissima* (van Kleunen et al. 2004).

We applied a 0.75 mM JA spray to randomly selected plants in each plot. (\pm)-jasmonic acid neat liquid concentrate (Sigma-Aldrich, Missouri, USA; CAS #: 77026-92-7) was mixed with an ethanol solvent and subsequently diluted with distilled water to produce the 0.75 mM JA solution. Treated plants were sprayed with the JA solution across the entire aboveground shoot to

runoff, i.e., the solution began dripping off the leaves (Thaler et al. 1996). Plants were sprayed at different timings (i.e., when the spray occurred on a plant; early, middle, and late timings in the experiment) and frequencies (i.e., how many times the spray occurred on a plant across those three timings) throughout the growing season (Figure 2.1b). When plants were not sprayed with JA, they were sprayed with water solution where the volume of JA was substituted with ethanol. At five time points after the first spray in the experiment, we measured the same response variables as in the mirid-feeding experiment.

Statistical analyses

We used generalized linear mixed models (GLMMs) to analyze the response variables from both experiments with the *lme4* package in R version 4.1.1 (Bates et al. 2015; R Core Team 2021). For the mirid-feeding experiment, each model contained predictors for observation time (days since feeding), mirid-feeding status (exposed/not exposed), and an interaction between time and mirid-feeding status. Since the same plants were measured repeatedly, plant was incorporated into the model as a random effect. We included another random effect for putative genotype based on the proximity of location rhizomes were collected from. Chewing herbivory and pathogen damage were calculated by multiplying the number of leaves of each plant by its estimated proportion of tissue lost to or covered with damage. Chewing herbivory and pathogen damage were log-transformed. Plants without visible damage had 0.5% added to their total to allow for log-transformations, which is biologically realistic given that most plants likely had some small amount of herbivory and pathogen damage that may not have been seen.

For the second experiment, plants were coded to account for the change in treatment status throughout time. For example, a plant that was sprayed once at the beginning of the experiment and again toward the later part of the experiment would be considered an early-only plant until it

was sprayed for the second time, which it was then considered an early + late plant. For that reason, the number of plants in any given spray timing group ranged from seven to 126 and the sample sizes of the groups changed with additional sprays. For a full list of sample sizes at each observation point, see Table S2.3. Three GLMMs per response variable were used to fit the data in this experiment and contained predictors for observation time, time and frequency of JA spray, and the interaction between observation time and the timing and frequency of JA spray (Figure 2.1b). The first model was fit only to the data from plants that did not receive exogenous JA (heretofore referred as unsprayed) and those that were sprayed with JA at the early timing (08-Jul-2020). The second model was fit to data for unsprayed plants, early JA sprays (08-Jul-2020), and middle JA sprays (30-Jul-2020). The third model was fit to data for unsprayed plants and plants with all three JA spray timings—early (08-Jul-2020), middle (30-Jul-2020), and late (23-Aug-2020). Chewing herbivory and pathogen damage values were calculated as described in the mirid-feeding experiment and log-transformed.

For both experiments, we calculated effect sizes at each observation date to determine the impact of different spray timings and frequencies. Confidence intervals (95% CI reported in parentheses after each effect size) for those predictions were determined using semi-parametric bootstrapping with the *bootMer* function with 2000 iterations in the *lme4* package (Bates et al. 2015). If the confidence intervals between two treatments did not overlap zero, the effect was deemed significant. We evaluated the significance of each model through the marginal R^2 (i.e., the variance explained by the fixed effects) and conditional R^2 (i.e., the variance explained by the fixed and random effects) with the *r.squaredGLMM* function from *MuMIn v1.40.0* (Bartón 2017).

The following R packages were used for data wrangling: *dplyr v1.0.8* (Wickham et al. 2022), *lubridate v1.7.10* (Grolemund and Wickham 2011), *googlesheets v1.0.0* (Bryan 2021), *tidyr*

v1.2.0 (Wickham and Girlich 2022), and *tibble* *v3.1.6* (Müller and Wickham 2021). Data was visualized with *ggplot2* *v3.3.5* (Wickham 2016), *ggpubr* (Kassambara 2020), and *ggbeeswarm* *v0.6.0* (Clarke and Sherrill-Mix 2017).

Results

How does a single early-season herbivory event alter community interactions and plant growth?

In this experiment, we evaluated how the difference between undamaged plants (control) and plants fed upon by mirids early in the growing season varied across time for each of the response variables. The effects of observation time, mirid-feeding, and their interaction explained 31.9%, 51.0%, and 26.5% of the variance (marginal R^2) in chewing herbivory, pathogen damage, and plant height, respectively (Table S2.4). The effect of mirid feeding on subsequent chewing herbivory varied through time (Figure 2.2a). On 18-Jul-2018, mirid-damaged plants had 27.3% (95% CI; 24.5-30.0%) lower chewing herbivory than control plants, yet in absolute terms, there was little chewing herbivory damage on plants in the first few weeks after transplant (e.g., the absolute mean difference between mirid-fed plants and control was only 0.2 leaves with chewing herbivory). However, by the last observation date (22-Sep-2019), plants with mirid damage received 37.6% (32.3-43.0%) higher chewing herbivory than control plants (absolute mean difference between mirid-fed plants and control was 1.7 leaves with chewing herbivory). Pathogen damage on mirid-fed plants exhibited a similar trajectory to chewing herbivory but only had significantly higher pathogen damage on the control plants at the final observation date (Figure 2.2b). Early in the experiment on 18-Jul-2019, *Slaterocoris* sp.-damaged plants had 56.0% (54.4-57.4%) less area covered in pathogen damage than plants without early mirid feeding. By the end of the observation period, the direction of that effect reversed; plants with mirid feeding having 7.1% (1.0-13.6%) higher pathogen damage than control plants. Finally, plant height showed a

small effect of mirid feeding (Fig 2c), with lower plant height in mirid-fed plants than unfed plants at all observations (range of mean effects across all four observation dates: 6.5-21.0% (5.1-36.5%) decrease). We found no difference in rhizome volume between mirid-fed and control plants (Figure S2.3). See Figure S2.4 for plotted effect sizes and Table S2.5 for a table of all effect sizes throughout time.

How does variation in the timing of herbivory change community interactions and plant growth?

To evaluate the community and plant effects of herbivory timing, we examined the differences in effect sizes between unsprayed and JA sprayed plants for the three response variables. While there are many pairwise comparisons to discuss in this experiment, we address the most interesting here, but see Figure S2.5 for plotted effect sizes and Table S2.7 for a table of all pairwise effect sizes.

The effects of chewing herbivory varied greatly between sprayed and unsprayed plants over time. The fixed effects of the early, mid, and late chewing herbivory models only explained 0.4%, 1.3%, and 2.9% (marginal R^2) of the total variance, respectively (Table S2.6). Early-only sprays resulted in consistently higher chewing herbivory at all observation dates compared to unsprayed plants. On 18-Jul-2020, early-only plants had 17.0% (5.6-29.8%) higher chewing herbivory damage than unsprayed plants (Figure 3a). On 09-Sep-2020, that difference remained roughly constant to 20.2% (5.8-36.5%). When compared with unsprayed plants, middle-only plants did not differ significantly (Figure 3b). Plants with a late-only spray had 34.2% (7.7-53.0%) lower chewing herbivory on 30-Aug-2020, but no significant difference on 09-Sep-2020 (Figure 3c).

We also found significant differences in chewing herbivory between the three spray timings. Middle-only plants had lower chewing herbivory than early-only plants at the first two

observation dates (14-Aug-2020: 36.9% (34.3-39.4%) decrease; 30-Aug-2020: 15.3% (12.7-17.7%) decrease; 09-Sep-2020, no significant difference; Fig 3b). Late-only plants had lower chewing herbivory than early-only and middle-only plants at both observation dates. On 30-Aug-2020, late-only sprays resulted in 35.2% (23.1-45.4%) and 25.7% (15.9-34.3%) lower chewing herbivory than early-only and middle-only plants, respectively (Figure 3c). On 09-Sep-2020, that effect remained, with late-only plants having 24.0% (11.0-35.0%) and 36.2% (16.9-34.4%) lower chewing herbivory than their early-only and middle-only spray counterparts, respectively (Figure 3c).

Plant surface area covered in pathogen damage followed a slightly different pattern than chewing herbivory. The fixed effects of the early, mid, and late models explained 7.1%, 3.0%, and 3.3% (marginal R^2) of the total variance in pathogen damage, respectively (S9). The effect of early-only spray dampened throughout time in comparison to unsprayed plants; on 18-Jul-2020, those plants had 12.4% (4.3-19.9%) lower pathogen damage than unsprayed plants, but by 30-Aug-2020, that difference was not apparent (5.6% (13.9% decrease-3.6% increase) decrease; Figure 2.4a). Middle-only plants did not have significantly different pathogen damage than unsprayed plants on any observation date (Figure 2.4b). Late-only plants were not significantly different from unsprayed plants, but they had lower pathogen damage than early-only and middle-only plants on 30-Aug-2020 (Figure 2.4c).

Plant height was significantly different when compared among spray and observation timings, though effect sizes were typically small. The early, mid, and late models respectively explained 11.6%, 0.7%, and 1.6% (marginal R^2) of the variance. Middle-only plants were 1.1% (0.5-1.6%) and 1.5% (0.9-1.9%) taller than early-only plants on 30-Aug-2020 and 09-Sep-2020, respectively (Figure 2.5b). Late-only plant heights were not different from early-only plant heights,

but these plants were shorter than middle-only plants regardless of observation date, with a 3.9% (1.0-7.7%) and 4.1% (1.1-7.9%) decrease on 30-Aug-2020 and 09-Sep-2020, respectively (Figure 2.5c).

How does the frequency of herbivory events change community interactions and plant growth?

By spraying plants either once or twice at different intervals throughout the summer, we were able to test for accumulating effects of multiple herbivory events on the response variables. Although many pairwise comparisons between treatments and observational timings are possible, we focus on those that illustrate different community and plant responses to the frequency of applications. See Table S2.7 for a table of all effect sizes and Figure S2.6 for plotted effect sizes.

First, early + middle spray treatment plants had significantly higher levels of chewing herbivory than unsprayed plants and middle-only plants at multiple observation dates, but they were never different from early-only plants (Figure 3b). On 14-Aug-2020, early + middle plants exhibited 76.7% (20.0-160.0%) and 96.7% (53.9-151.5%) greater chewing herbivory when compared to unsprayed and middle-only plants, respectively. However, chewing herbivory on early + middle plants were not significantly different from unsprayed plants by 30-Aug-2020, yet still they maintained 34.0% (10.7-62.3%) higher chewing herbivory than middle-only plants at that date. These plants also experienced higher chewing herbivory when compared to late-only plants on both 30-Aug-2020 and 09-Sep-2020 (Figure 3c).

Regardless of observation date, plants treated early + late experienced significantly lower chewing herbivory than single spray or unsprayed plants (Figure 3c). For example, plants with early + late sprays had 34.1% (2.9-55.3%) and 48.8% (23.9-65.6%) lower chewing herbivory than unsprayed plants on 30-Aug-2020 and 09-Sep-2020, respectively. The effect of middle + late sprays varied by observation date. On 30-Aug-2020, these plants were only different from late-

only plants, with 48.7% (42.3-55.2%) higher chewing herbivory. However, by 09-Sep-2020, middle + late plants had significantly lower chewing herbivory than unsprayed, early-only, middle-only, and late-only plants (Figure 3c).

Pathogen damage exhibited a different pattern of response compared with those of chewing herbivory in response to the frequency of JA sprays. Early + middle plants had significantly lower pathogen damage than other plants toward the end of the season (Figs. 4b-c). On 09-Sep-2020, early + middle plants had 43.2% (16.4-61.4%) lower pathogen damage compared to unsprayed plants, 37.1% (20.8-50.0%) lower damage than early-only plants, 38.6% (25.9-49.2%) lower damage compared to middle-only plants, and 43.6% (38.0-48.6%) lower damage compared to late-only plants. Early + late treated plants had significantly lower pathogen damage than unsprayed and all single-spray plants on 30-Aug-2020, but were only significantly different from the two single-spray treatments by 09-Sep-2020, with a 18.6% (5.0-30.3%) decrease in pathogen damage against middle-only treated plants and a 19.5% (14.8-24.0) decrease against late-only treated plants (Figure 2.4c). Conversely, when compared to unsprayed or single spray plants, middle + late treated plants only differed from late-only treated plants on 30-Aug-2020 and not with any single spray timing on 09-Sep-2020.

Plant height was different from other treatments only in those plants receiving two sprays that included a late spray, and these effects varied by observation date. Of the double-spray plants, only middle + late treated plants grew to different heights than unsprayed plants (Figure 2.5c). Early + late plants were taller than late-only plants on 30-Aug-2020 (1.6% (0.4-2.5%) increase) and 09-Sep-2020 (2.0% (0.8-2.8%) increase). Middle + late treated plants were significantly shorter than middle-only plants on 30-Aug-2020 (12.0% (7.1-18.3%) decrease) and 09-Sep-2020

(11.6% (6.8-17.8%) decrease) and late-only plants on Aug 30 (8.4% (6.2-11.4%) decrease) and 09-Sep-2020 (7.9% (5.7-10.8%) decrease).

Finally, we can also compare how different double spray timings affect these response variables. For example, on 30-Aug-2020, middle + late treated plants had 16% (12.8-19.0%) lower chewing herbivory than early + middle treated plants, but they also had 48.6% (47.6-49.6%) higher chewing herbivory than early + late treated plants (Figure 3c). However, on that same date, middle + late treated plants had no difference in chewing herbivory from unsprayed, early-only, and middle-only treated plants.

Discussion

Using varying timings of treatments and measurements, we showed that both timing and frequency of herbivory along plant ontogeny cannot be overlooked when measuring community interactions and plant growth. Likewise, observation date directly impacts the interpretation of effect sizes between plants with and without herbivory because plant responses are dynamic and vary with the amount of time since their encounter with herbivores. These results indicate that antagonism at different moments during a plant's life cycle can cause different defense and growth responses, with cascading effects on species interactions (Walck et al. 1999).

To interpret the temporally variable effects of herbivory, we should account for the phenology of *S. altissima*, especially when it may invest more in vegetative growth, defense, asexual reproduction, and sexual reproduction. Early in the season, when we applied *Slaterocoris* sp. to feed on the plants, *S. altissima* was actively growing and investing largely in aboveground vegetative growth (Walck et al. 1999). In that time, growing taller represents a trade-off while goldenrod is also competing for light resources and reducing apparency to key herbivores, such as the goldenrod ball gallfly, *E. solidaginis*. Relative to control plants, *Slaterocoris* sp.-treated plants

grew shortest early the experiment (Figure 2.2c), which follows a similar trajectory as an earlier study investigating early-season defoliation's effect on height (Meyer 1998).

Reducing apparency to other herbivores or investing resources in other defensive strategies rather than growth are two possible explanations for this response. Our results indicate the latter explanation is plausible, given the reduced chewing herbivory and pathogen damage early in the experiment. However, the early induction of these plants was short-lived since less than two months later, the direction of damage patterns reversed, and the previously induced plants were more susceptible to damage at the end of the season. In September, *S. altissima* plants are typically investing in the reproductive resources of flowers and seeds (Walck et al. 1999), which may leave them more vulnerable to attack, and at this time their grasshopper generalist chewing herbivores are most abundant. By then, pathogens such as leaf rust may have dispersed between plants and are spreading within plants. This induced defense followed by induced susceptibility has been noted in other systems and may be associated with the costs of defending near the time of the herbivory event (Underwood 1998). Evaluating the time-dependent changes in phytochemical and antagonist preferences in plant tissue is a key next step in identifying the mechanisms behind these patterns.

Traits in plant-insect ecology are often measured statically, yet that is not how they mediate interactions in nature. For example, we measured rhizome volume in the winter following our mirid-feeding experiment and found no difference between control and mirid-fed plants (Figure S2.3). Due to the destructive nature of rhizome collection, we were unable to determine if these patterns held into the early-spring sprouting of the following year. In the JA-spray experiment, the results demonstrated that community-level responses to a simulated plant-herbivore encounter are timing-dependent; when compared to unsprayed plants, early and late

season sprays resulted in significant, yet opposite, effects on subsequent herbivory damage, while mid-season sprays compared to unsprayed plants were not statistically significant. Perhaps this lack of effect is caused by investment in growth or reproduction or that similar stimulations to those from a JA spray are uncommon at this time in the growing season. Therefore, outstanding questions are how the timing of plant defensive response is an evolutionary response to the timing of herbivore arrival and how reproductive responses may be undetectable due to measurement timing.

Intuitively, we may expect multiple stressor effects to accumulate or interact over time, but our results do not suggest that is always the case. As in the mirid feeding experiment, JA sprays had significant effects on the plants' interactions with its antagonist community. Early-only spray plants experienced elevated chewing herbivore damage at all observation times (Figure 2.3a), yet for early + middle plants, this positive effect dissipated as the season progressed (Figure 2.3b). This result suggests that even if plants are antagonized early in the season, another antagonistic event does not predestine them for more chewing herbivory in the future. However, the opposite is observed for this comparison and pathogen damage; early-only plants received (relative to controls) more pathogen damage earlier in the season than later in the season (Figure 2.4a), but plants with early + middle treatments received relatively the most pathogen damage later in the season (Figure 2.4b). This result presents a possible trade-off in resistance to herbivores and pathogens, which has been observed with mixed support across various plant stressor events (Biere et al. 2004; Löser et al. 2021). Additionally, the magnitude of these differences in height between single-spray and double-spray plants may affect apparency to herbivores and pollinators, with important implications for competition and fitness. Again, the contrasting effects we observed with different spray timings, spray frequency, and observation timings relative to the study organism's

phenology and ontogeny necessitate greater emphasis on the temporal context of ecological experiments.

Although repeated measurements can reveal important patterns in temporal ecology, it can make it more difficult to identify the mechanisms behind the trends. Repeated measurements require live organisms at each observation date and thus prohibits the destructive collection of biological material for chemical analyses, as well as certain measurements of fitness and growth such as rhizome collection and dry biomass. Without some of these direct measures, it is difficult to ascertain why pathogens did not spread within plants in some treatments or why chewing herbivores may have avoided insect-fed plants early in the experiment. Future field work could pair plants to apply herbivory with repeated observations throughout time, with some plants serving to measure community interactions and other plants to measure underlying chemical changes, which has been achieved at single time observations already (González et al. 2018).

The biological patterns we observed reflect both treatment and observation timing, and these results highlight how important it is for temporal context to be explicitly incorporated into ecological experiments. Like many other long-lived plant species, *S. altissima* is a perennial forb whose genets interact with many species at inter-annual and intra-annual time scales. We should expect that in systems containing these plant-herbivore-pathogen interactions, responses to ecological events are scale-dependent and hierarchical. By repeatedly measuring communities at moments of time that eventually build up to the long-term trajectories, we can adjust our ecological interpretations and aim for a more nuanced, longitudinal understanding of nature in a changing world.

REFERENCES

- Agrawal AA (1998) Induced responses to herbivory and increased plant performance. *Science* 279:1201–1202. <https://doi.org/10.1126/science.279.5354.1201>
- Anderson SS, McCrea KD, Abrahamson WG, Hartzel LM (1989) Host genotype choice by the ball gallmaker *Eurosta solidaginis* (Diptera: Tephritidae). *Ecology* 70:1048–1054. <https://doi.org/10.2307/1941373>
- Babst BA, Ferrieri RA, Gray DW, et al (2005) Jasmonic acid induces rapid changes in carbon transport and partitioning in *Populus*. *New Phytologist* 167:63–72. <https://doi.org/10.1111/j.1469-8137.2005.01388.x>
- Bahlai CA, Hart C, Kavanaugh MT, et al (2021) Cascading effects: insights from the U.S. Long Term Ecological Research Network. *Ecosphere* 12:e03430. <https://doi.org/10.1002/ecs2.3430>
- Bartón K (2017) MuMin: Multi-model inference. <https://cran.r-project.org/web/packages/MuMIn/index.html>
- Barton KE (2013) Ontogenetic patterns in the mechanisms of tolerance to herbivory in *Plantago*. *Annals of Botany* 112:711–720. <https://doi.org/10.1093/aob/mct083>
- Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48. <https://doi.org/10.18637/jss.v067.i01>
- Biere A, Marak HB, van Damme JMM (2004) Plant chemical defense against herbivores and pathogens: generalized defense or trade-offs? *Oecologia* 140:430–441. <https://doi.org/10.1007/s00442-004-1603-6>
- Blitzer EJ, Welter SC (2011) Emergence asynchrony between herbivores leads to apparent competition in the field. *Ecology* 92:2020–2026. <https://doi.org/10.1890/11-0117.1>
- Borer ET, Seabloom EW, Gruner DS, et al (2014) Herbivores and nutrients control grassland plant diversity via light limitation. *Nature* 508:517–520. <https://doi.org/10.1038/nature13144>
- Bryan J (2021) googlesheets4: Access Google Sheets using the Sheets API V4. <https://googlesheets4.tidyverse.org>
- Burghardt KT (2016) Nutrient supply alters goldenrod's induced response to herbivory. *Functional Ecology* 30:1769–1778. <https://doi.org/10.1111/1365-2435.12681>
- Cain ML (1990) Models of clonal growth in *Solidago altissima*. *The Journal of Ecology* 78:27. <https://doi.org/10.2307/2261034>

- Clark GF, Johnston EL (2011) Temporal change in the diversity–invasibility relationship in the presence of a disturbance regime. *Ecology Letters* 14:52–57. <https://doi.org/10.1111/j.1461-0248.2010.01550.x>
- Clarke E, Sherrill-Mix S (2017) ggbeeswarm: Categorical scatter (violin point) plots. <https://github.com/eclarke/ggbeeswarm>
- Cronin JT (2007) Shared parasitoids in a metacommunity: indirect interactions inhibit herbivore membership in local communities. *Ecology* 88:2977–2990. <https://doi.org/10.1890/07-0253.1>
- Ekholm A, Tack AJM, Pulkkinen P, Roslin T (2020) Host plant phenology, insect outbreaks and herbivore communities – the importance of timing. *Journal of Animal Ecology* 89:829–841. <https://doi.org/10.1111/1365-2656.13151>
- Fukami T (2015) Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annu Rev Ecol Evol Syst* 46:1–23. <https://doi.org/10.1146/annurev-ecolsys-110411-160340>
- González JB, Petipas RH, Franken O, et al (2018) Herbivore removal reduces influence of arbuscular mycorrhizal fungi on plant growth and tolerance in an East African savanna. *Oecologia* 187:123–133. <https://doi.org/10.1007/s00442-018-4124-4>
- Grolemund G, Wickham H (2011) Dates and times made easy with lubridate. *Journal of Statistical Software* 40:1–25. <https://doi.org/10.18637/jss.v040.i03>
- Helms AM, Moraes CMD, Tooker JF, Mescher MC (2013) Exposure of *Solidago altissima* plants to volatile emissions of an insect antagonist (*Eurosta solidaginis*) deters subsequent herbivory. *PNAS* 110:199–204. <https://doi.org/10.1073/pnas.1218606110>
- Howe GA, Jander G (2008) Plant immunity to insect herbivores. *Annu Rev Plant Biol* 59:41–66. <https://doi.org/10.1146/annurev.arplant.59.032607.092825>
- Jackson MC, Pawar S, Woodward G (2021) The temporal dynamics of multiple stressor effects: from individuals to ecosystems. *Trends in Ecology & Evolution* 36:402–410. <https://doi.org/10.1016/j.tree.2021.01.005>
- Kalske A, Shiojiri K, Uesugi A, et al (2019) Insect herbivory selects for volatile-mediated plant-plant communication. *Current Biology* 29:3128–3133.e3. <https://doi.org/10.1016/j.cub.2019.08.011>
- Kant MR, Jonckheere W, Knecht B, et al (2015) Mechanisms and ecological consequences of plant defence induction and suppression in herbivore communities. *Annals of Botany* 115:1015–1051. <https://doi.org/10.1093/aob/mcv054>
- Karban R (2020) The ecology and evolution of induced responses to herbivory and how plants perceive risk. *Ecological Entomology* 45:1–9. <https://doi.org/10.1111/een.12771>

- Kassambara A (2020) ggpubr: “ggplot2” Based Publication Ready Plots.
<https://github.com/kassambara/ggpubr>
- Löser TB, Mescher MC, De Moraes CM, Maurhofer M (2021) Effects of root-colonizing fluorescent *Pseudomonas* strains on *Arabidopsis* resistance to a pathogen and an herbivore. *Appl Environ Microbiol* 87:e02831-20. <https://doi.org/10.1128/AEM.02831-20>
- Meyer GA (1998) Pattern of defoliation and its effect on photosynthesis and growth of goldenrod. *Functional Ecology* 12:270–279. <https://doi.org/10.1046/j.1365-2435.1998.00193.x>
- Müller K, Wickham H (2021) tibble: Simple data frames. <https://tibble.tidyverse.org/>
- Nykanen H, Koricheva J (2004) Damage-induced changes in woody plants and their effects on insect herbivore performance: a meta-analysis. *Oikos* 104:247–268.
<https://doi.org/10.1111/j.0030-1299.2004.12768.x>
- Ochoa-López S, Villamil N, Zedillo-Avelleyra P, Boege K (2015) Plant defence as a complex and changing phenotype throughout ontogeny. *Ann Bot* 116:797–806.
<https://doi.org/10.1093/aob/mcv113>
- Pang W, Crow WT, Luc JE, et al (2011) Comparison of water displacement and WinRHIZO software for plant root parameter assessment. *Plant Disease* 95:1308–1310.
<https://doi.org/10.1094/PDIS-01-11-0026>
- R Core Team (2021) R: A language and environment for statistical computing
- Rasmussen NL, Yang LH (2022) Timing of a plant–herbivore interaction alters plant growth and reproduction. *Ecology* e3854. <https://doi.org/10.1002/ecy.3854>
- Rusman Q, Lucas-Barbosa D, Hassan K, Poelman EH (2020) Plant ontogeny determines strength and associated plant fitness consequences of plant-mediated interactions between herbivores and flower visitors. *Journal of Ecology* 108:1046–1060.
<https://doi.org/10.1111/1365-2745.13370>
- Ryo M, Aguilar-Trigueros CA, Pinek L, et al (2019) Basic principles of temporal dynamics. *Trends in Ecology & Evolution* 34:723–733. <https://doi.org/10.1016/j.tree.2019.03.007>
- Thaler JS, Stout MJ, Karban R, Duffey SS (1996) Exogenous jasmonates simulate insect wounding in tomato plants (*Lycopersicon esculentum*) in the laboratory and field. *J Chem Ecol* 22:1767–1781. <https://doi.org/10.1007/BF02028503>
- Tooker JF, De Moraes CM (2009) A gall-inducing caterpillar species increases essential fatty acid content of its host plant without concomitant increases in phytohormone levels. *MPMI* 22:551–559. <https://doi.org/10.1094/MPMI-22-5-0551>

- Underwood NC (1998) The timing of induced resistance and induced susceptibility in the soybean-Mexican bean beetle system. *Oecologia* 114:376–381. <https://doi.org/10.1007/s004420050460>
- van Dijk LJA, Ehrlén J, Tack AJM (2020) The timing and asymmetry of plant–pathogen–insect interactions. *Proceedings of the Royal Society B: Biological Sciences* 287:20201303. <https://doi.org/10.1098/rspb.2020.1303>
- van Kleunen M, Ramponi G, Schmid B (2004) Effects of herbivory simulated by clipping and jasmonic acid on *Solidago canadensis*. *Basic and Applied Ecology* 5:173–181. <https://doi.org/10.1078/1439-1791-00225>
- Walck JL, Baskin JM, Baskin CC (1999) Relative competitive abilities and growth characteristics of a narrowly endemic and a geographically widespread *Solidago* species (Asteraceae). *American Journal of Botany* 86:820–828. <https://doi.org/10.2307/2656703>
- Wang M, Bezemer TM, van der Putten WH, et al (2018) Plant responses to variable timing of aboveground clipping and belowground herbivory depend on plant age. *Journal of Plant Ecology* 11:696–708. <https://doi.org/10.1093/jpe/rtx043>
- Wauchope HS, Amano T, Geldmann J, et al (2021) Evaluating impact using time-series data. *Trends in Ecology & Evolution* 36:196–205. <https://doi.org/10.1016/j.tree.2020.11.001>
- Wickham H (2016) *Elegant graphics for data analysis*. Springer-Verlag New York
- Wickham H, François R, Henry L, Müller K (2022) *dplyr: a grammar of data manipulation*. <https://dplyr.tidyverse.org>
- Wickham H, Girlich M (2022) *tidyr: tidy messy data*. <https://tidyr.tidyverse.org>
- Wise MJ, Abrahamson WG (2008) Ducking as a means of resistance to herbivory in tall goldenrod, *Solidago altissima*. *Ecology* 89:3275–3281
- Yang LH (2020) Toward a more temporally explicit framework for community ecology. *Ecological Research* 35:445–462. <https://doi.org/10.1111/1440-1703.12099>
- Yang LH, Censer ML (2020) Seasonal windows of opportunity in milkweed–monarch interactions. *Ecology* 101:e02880. <https://doi.org/10.1002/ecy.2880>
- Yang LH, Censer ML, Morgan LJ, Hall GW (2020) Species-specific, age-varying plant traits affect herbivore growth and survival. *Ecology* 101:e03029. <https://doi.org/10.1002/ecy.3029>

FIGURES

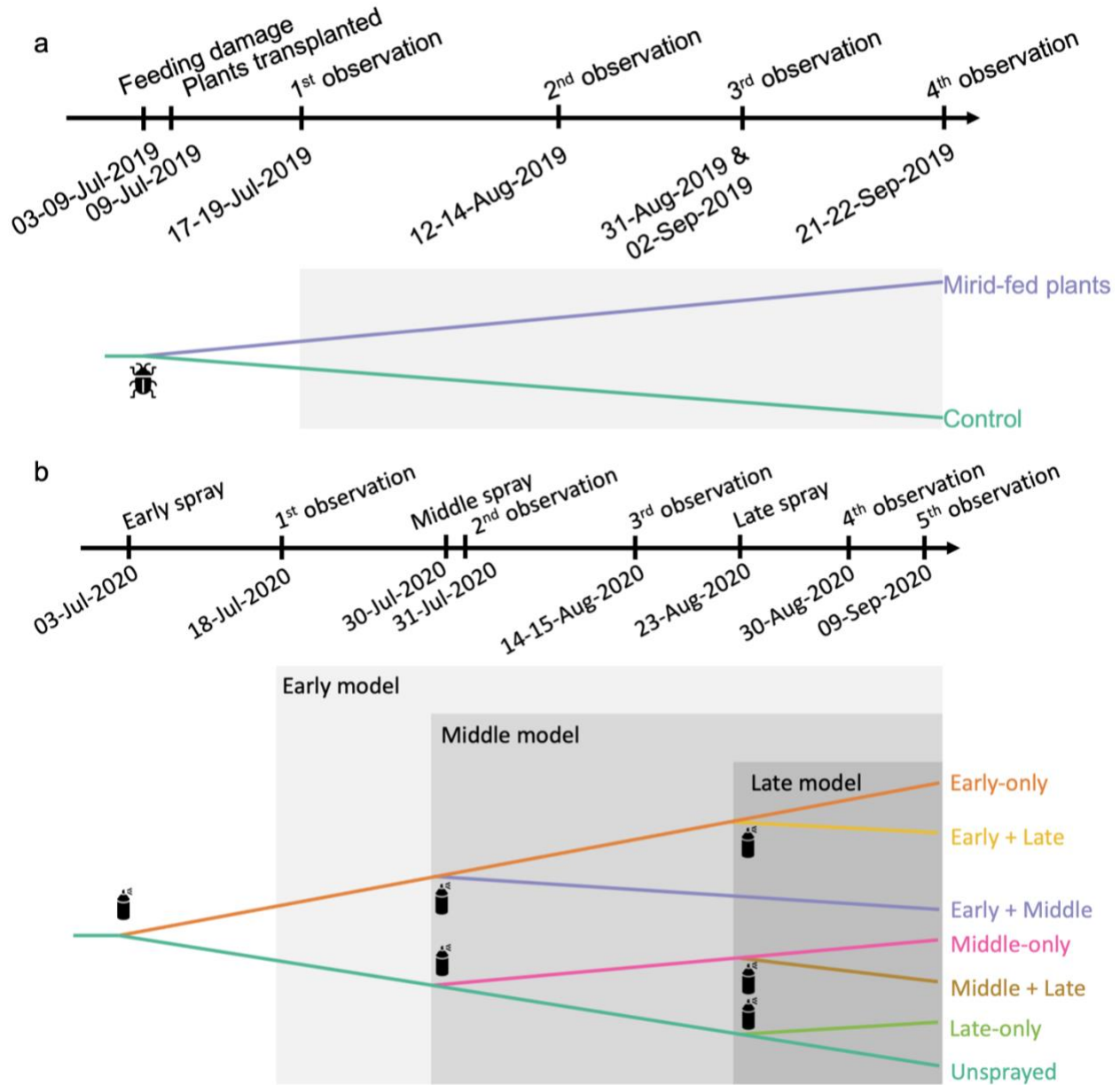


Figure 2.1 a) Experimental setup of the early-season herbivory experiment with *Slaterocoris* sp. and b) the varying timing and frequency JA spray experiment. The start of each diverging colored line indicates when herbivory (mirid feeding or JA spray) was applied for each group of *S. altissima* individuals. The gray boxes are aligned with the timeline to indicate the observation dates covered in each model.

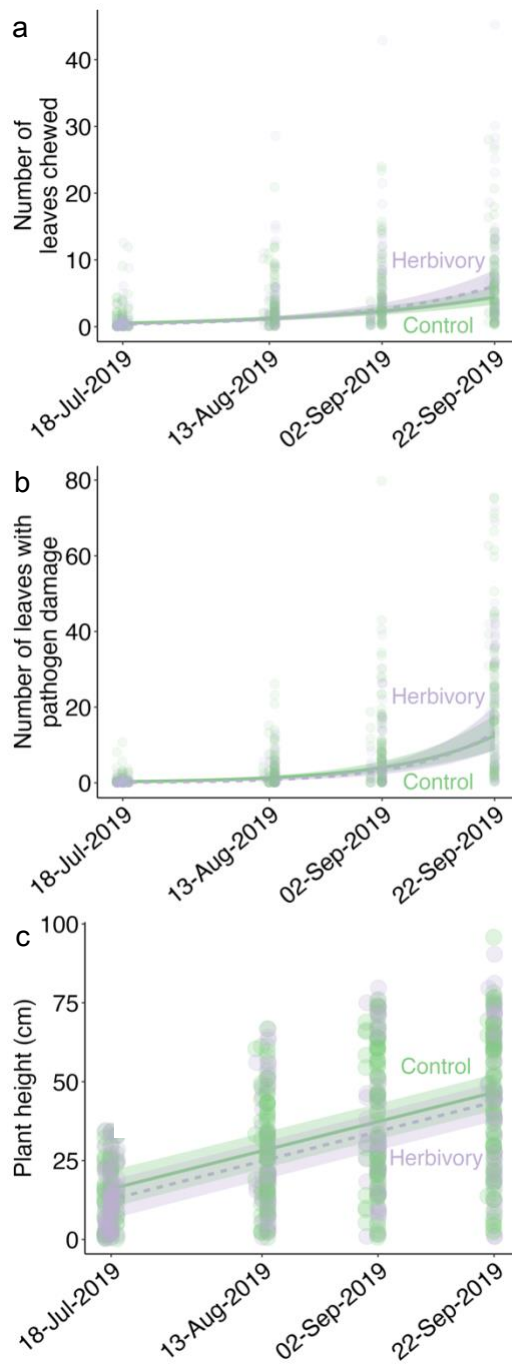


Figure 2.2 Mirid-feeding experiment model predictions and 95% CIs by control (green, solid line) and mirid-fed (purple, dashed line) plants across time for a) chewing herbivory, b) pathogen damage, and c) plant height.

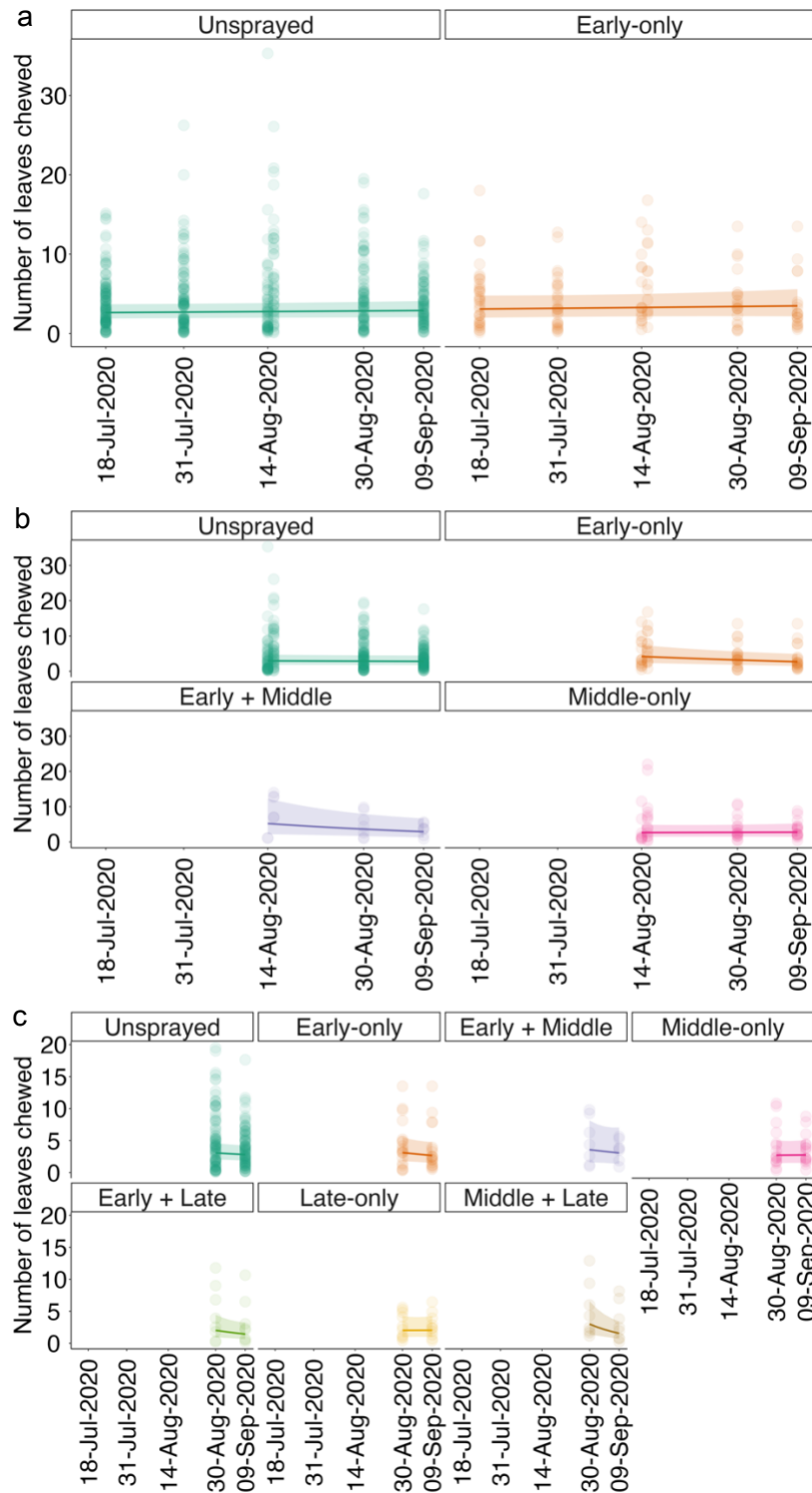


Figure 2.3 JA spray model predictions and 95% CIs for chewing herbivory for a) early treated plants, b) middle treated plants, and c) late treated plants.

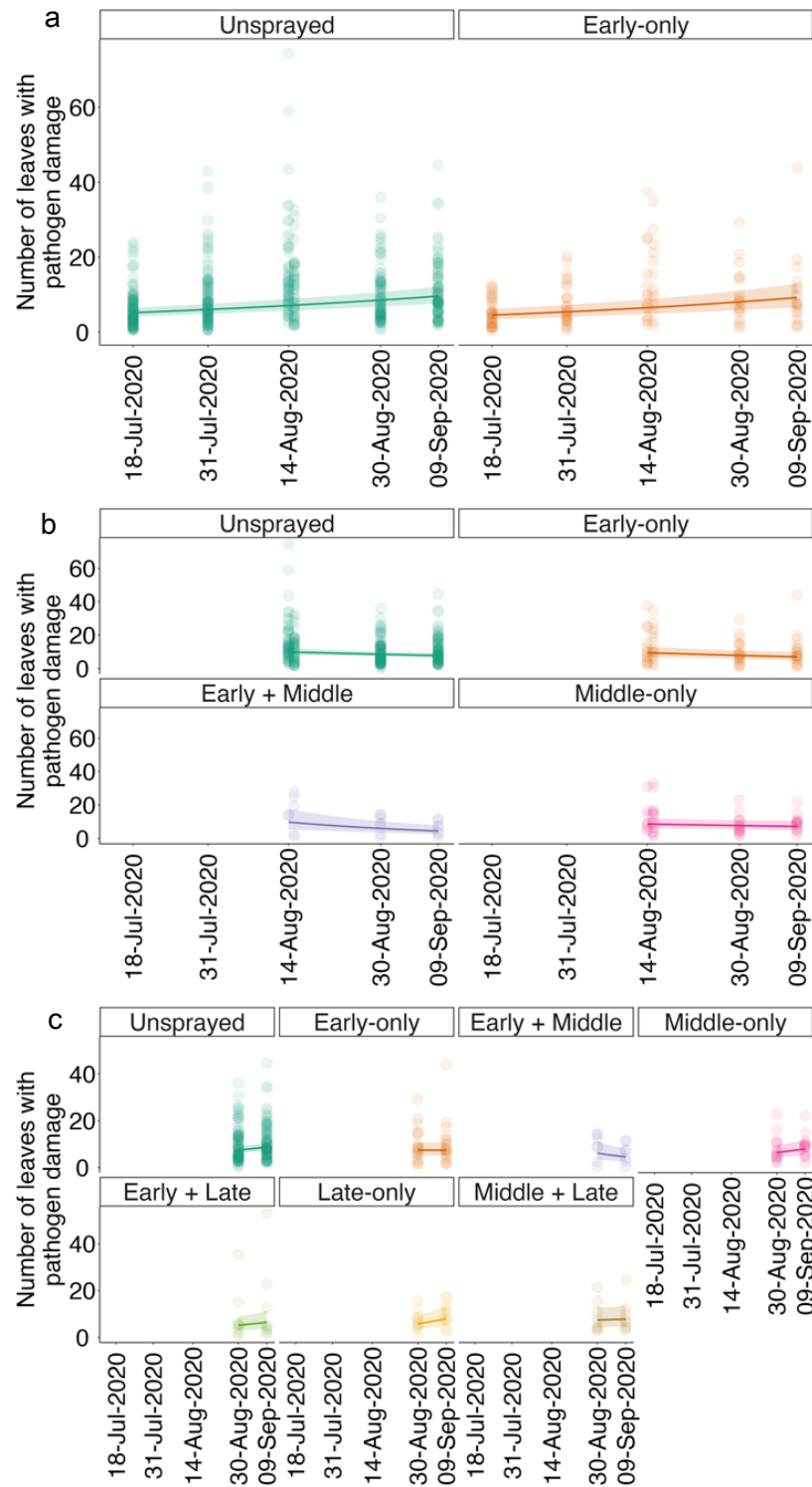


Figure 2.4 JA spray model predictions and 95% CIs for pathogen damage for a) early treated plants, b) middle treated plants, and c) late treated plants.

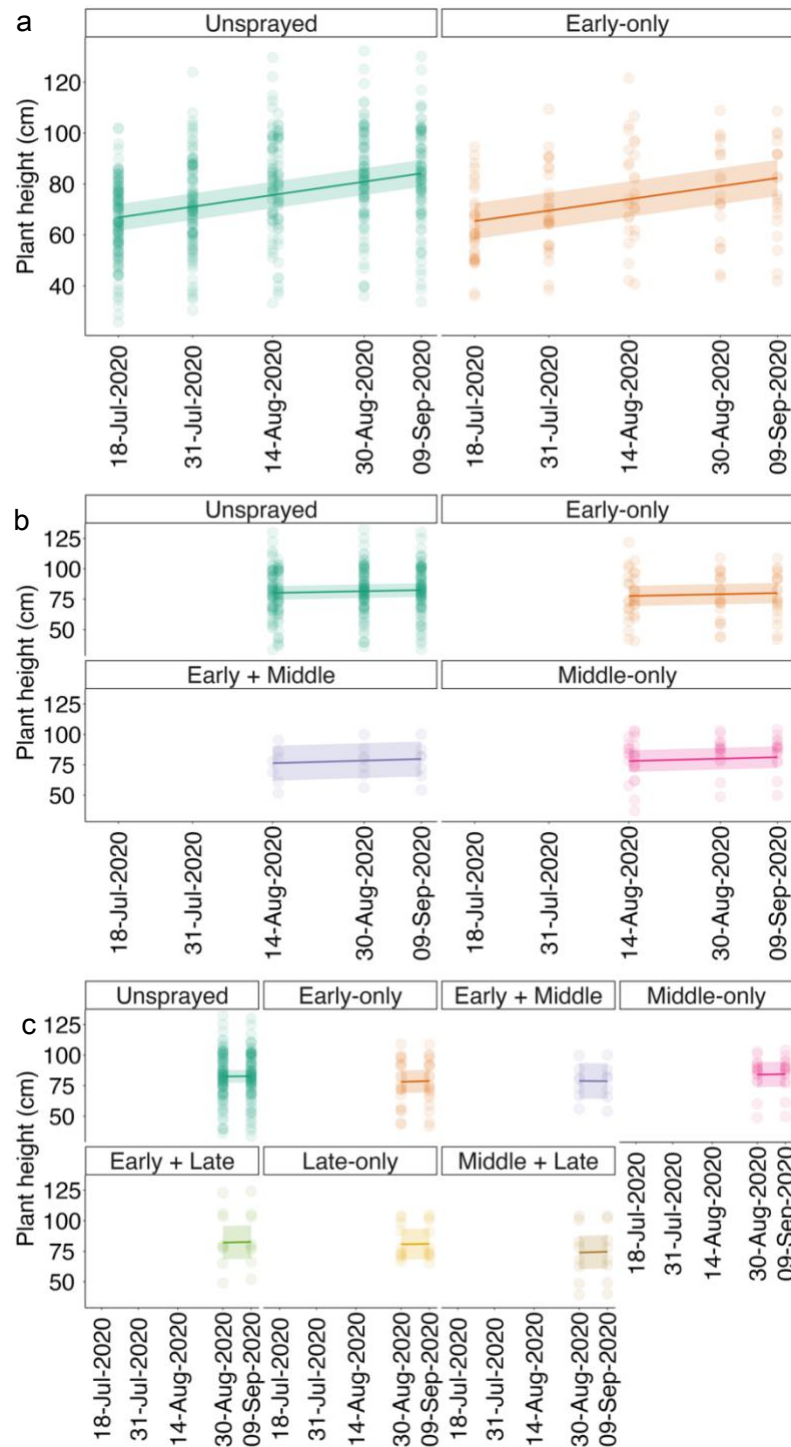


Figure 2.5 JA spray model predictions and 95% CIs for plant height for a) early treated plants, b) middle treated plants, and c) late treated plants.

CHAPTER 3:
PLANT TERPENOID DIVERSITY VARIES WITH TEMPERATURE,
PRECIPITATION, AND PHYLOGENY: A META-ANALYSIS

Abstract

Monoterpenoids and sesquiterpenoids are ubiquitous and diverse plant chemicals with ecological and human significance. Their availability and diversity within plants may be influenced by temperature, precipitation, and phylogeny. Plants produce diverse profiles of monoterpenoids and sesquiterpenoids to form mutualisms, defend against enemies, and protect against abiotic stress, but broad scale patterns in that diversity have not been investigated thoroughly. I conducted a meta-analysis using a global database of published literature to obtain the monoterpenoid and sesquiterpenoid profiles of 206 plant species across 277 sites. I identified increasing compound richness with increasing annual precipitation, as well as significant relationships in how terpene identity and structure vary with annual mean temperature, precipitation, and phylogenetic distance. Understanding how these patterns scale across various metrics of diversity illuminated the relative importance of abiotic factors, evolutionary history, and ecological interactions in plant production of lighter weight terpenoids, with implications for chemical ecology and applications across human industries.

Introduction

Biochemicals are the building blocks of plants that nourish and heal us, provide food and shelter for wildlife, and sequester carbon and pollutants. Functioning beyond plants' primary metabolic processes (e.g., photosynthesis), phytochemicals known as secondary metabolites were considered “waste” only a few decades ago (Hartmann 2007). While people applied these natural products in daily life, their diversity and distribution were still mysteries (Gershenzon and Dudareva 2007; Kessler and Kalske 2018). Two subclasses of secondary metabolites have been crucial to plant chemical ecology and evolution research—monoterpenoids and sesquiterpenoids. Thousands of these compounds have now been identified, and defining their ecological functions, evolution, and syntheses is an active area of investigation (Tholl 2015; Pichersky and Raguso 2018). Their wide array of forms and functions poses a central question—how does terpenoid diversity vary across abiotic conditions and evolutionary history? Collating 148 studies from a database of published plant terpenoids, in this study I conducted a meta-analysis of macroscale relationships between climate, phylogeny, and diversity in monoterpenoids and sesquiterpenoids to develop a foundational understanding about how these compounds vary in nature.

Phytochemical diversity can have profound impacts on differences in the ecological interactions between closely related species and of genotypes within species (Dyer *et al.* 2018, Philbin *et al.* 2022, Salazar and Marquis 2022). Throughout plant evolution, the retention of existing phytochemicals in a profile and the production of novel phytochemicals are the result of numerous biotic and abiotic selective pressures. Compound signals, such as terpenoid volatiles, are often received and effective at high specificity (Chen and Song 2008). However, a single terpenoid can also serve multiple functions (Pichersky and Raguso 2018). Therefore, producing

more compounds may allow plants to interact with more organisms across mutualistic and antagonistic capacities. Some compounds may also only be effective ecologically in tandem with other compounds (Richards *et al.* 2015), so phytochemical diversity *per se* may be advantageous. Similar to foundational work on identifying biogeographical patterns in plant species and trait diversity, we can begin macroscale studies in phytochemical diversity by describing its associations with evolutionary history and the abiotic environment.

Several experiments have tested hypotheses about phytochemical diversity primarily within the context of plants' ecological interactions and evolutionary history (Becerra 1997, Junker *et al.* 2017, Salazar *et al.* 2018, Whitehead *et al.* 2021). A longstanding hypothesis, the coevolutionary arms race between plants and insects, has presented phytochemical diversity as a function of insect herbivory (Ehrlich and Raven 1964). Plants are suggested to evolve more novel and “complex” compounds to defend against their enemies, while insects can evolve mechanisms to resist these defenses, such as toxin sequestration or the coopting of chemical cues to locate nutrient resources (Berenbaum and Feeny 1981, Després *et al.* 2007). Two outcomes may emerge in the relationship between evolutionary history and terpenoid diversity: (a) plants that are more closely related may have more similar terpenoid profiles since they experience similar ecological interactions or (b) plants that are more closely related have more different terpenoid profiles to differentiate themselves chemically from relatives. Still, defense is only one function of these compounds, and plants also communicate to pollinators and attract dispersers via terpenoids leading to even more species interaction-based hypotheses (Adler 2000). Most hypotheses about plant secondary metabolite diversity suggest adaptation to biotic conditions (e.g., *interaction diversity hypothesis*, *synergy hypothesis*; Whitehead *et al.* 2021). Testing of hypotheses addressing abiotic, or both abiotic and biotic, factors have relatively lagged behind

hypotheses based on antagonistic interactions with herbivores (Moreira *et al.* 2015, Abdala-Roberts *et al.* 2016).

Incorporating the abiotic environment, such as climate, nutrients, and light availability, into our understanding of phytochemical diversity has identified key patterns (Glassmire *et al.* 2019, Defossez *et al.* 2021). For example, annual mean temperature (temperature) is likely associated with phytochemical diversity given the effect of temperature on compound volatility, temperature-dependent enzymatic activity for chemical synthesis, and terpenoids' role in plant tolerance to thermal stress (Loreto and Schnitzler 2010). Compound volatility is determined by atmospheric pressure and temperature and is specific to each compound based on substructural composition. For instance, sesquiterpenoids, which have five more carbon atoms than monoterpenoids, and cyclical monoterpenoids may have more complex and diverse chemical substructures than some acyclic monoterpenoids, possibly decreasing their volatility. We may then expect plants to produce more structurally diverse terpenoid profiles in hotter climates, a prediction which has been previously supported in field observations where monoterpenoid and sesquiterpenoid richness was greater at sites with higher temperatures (Llusiá and Peñuelas 2000).

The level of annual precipitation may also influence terpenoid diversity indirectly through stomatal adaptations to avoid desiccation in dry regions. Terpenoids are often released through the stomata, along with gas-phase diffusion at the tissue-air surface (Niinemets *et al.* 2004). Stomatal conductance (e.g., opening of the stomata) is related to volatile terpenoid release through the diffusion gradient between the internal tissue and external environment (Fall and Monson 1992). The emission of those compounds with high solubility inside plant tissue, such as alcohols and aldehydes, is correlated with stomatal conductance (Niinemets *et al.* 2003). Earlier

studies found little effect of increased precipitation on the richness and emissions of monoterpenoids and sesquiterpenoids (Llusiá *et al.* 2010). However, measuring richness and emissions does not account for structural differences between compounds, which may exhibit different responses. For example, measures of richness are agnostic to the presence/absence of substructures that may make compounds more soluble in plant tissue. Therefore, we may expect that the terpenoid structural diversity increases with increasing precipitation through its indirect effect on stomatal conductance, and plants growing in more similar precipitation environments may be expected to produce more similar terpenoids.

The underpinnings of phytochemical diversity can be challenging to identify when many compounds arise from different synthetic pathways or require different laboratory methods for identification. Monoterpenoids and sesquiterpenoids can serve as a model set of compounds because they share biosynthetic pathways and are easily detected through methods such as gas chromatography-mass spectrometry (GC-MS). The 5-carbon precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) are universal in the synthesis of all plant terpenoids (Christianson 2008). One or two IPP molecules are condensed with DMAPP to form the next precursors for monoterpenoids or sesquiterpenoids, respectively. Coded in 15-30 genes, synthases then convert those molecules into a diverse array of compounds, followed by possible secondary reactions (Chen *et al.* 2011; Tholl 2006). In volatile and semi-volatile phases, terpenoids are perceived, often with high specialization, by antagonistic and mutualistic insects through olfactory organs, as revealed by coupled GC-electroantennographic detection (Bleeker *et al.* 2009). Terpenoids can make forests more flammable (Alessio *et al.* 2008), alarm plant neighborhoods of herbivore attack (Kigathi *et al.* 2019), and glow in the hazy mornings of the Blue Ridge Mountains (Went 1960). Some compounds also are used in human medicine for their

antioxidant, anti-viral, anti-bacterial and anti-inflammatory properties (Tetali 2019), and many are the basis of perfumes and fragrances (Schwab *et al.* 2013).

We can analyze phytochemical diversity through metrics typically employed for ecological diversity, but the many available metrics can lead to varying conclusions (Dyer 2018). If species richness is the count of species in a community or location, phytochemicals can also be tallied across various taxonomic and spatial scales to calculate compound richness. Similarly, as species can be deconstructed into functional units or traits (i.e., functional diversity), chemical compounds can also be parsed into their constituent parts, known as substructures, to calculate structural diversity. These chemical substructures, like 4-carbon rings, hydroxyl groups, and the location of double bonds between carbon molecules, are essential to the receptivity and volatility of terpenoids and may or may not be tied to ecological function. Here, α -diversity is the diversity (e.g., richness, evenness) within a sampling unit (i.e., a collection of plant tissue from one species at a well-described location), while β -diversity is the turnover between two sampling units. Changes in α -diversity along climate and phylogenetic gradients do not account for differences in the composition of the compounds or substructures in the terpenoid profile, whereas β -diversity does. Evaluating patterns α -diversity can describe the range of chemical diversity that a plant species or sample can hold, and studying β -diversity can explain how plant species or samples differ from each other chemically. To explore global terpenoid patterns within this diversity metrics framework, I conducted a meta-analysis to answer the following three questions:

1. Are plant monoterpene and sesquiterpene profiles more diverse in warmer and wetter climates?

2. Do plants growing in more similar climates have more similar monoterpenoids and sesquiterpenoids?
3. Do more closely related plant species have more similar monoterpenoids and sesquiterpenoids?

Methods

Data collection and software used

To develop a comprehensive database of literature of mono- and sesquiterpenoids, I searched for a broad array of plant terpenoid studies using combinations of the following terms on ISI Web of Science: terpene*, gc*ms, gc-ms, terpenoid*, ecolog*, and herbivor*. These searches resulted in 5471 unique peer-reviewed articles. Inclusion criteria were (1) sampled living plant tissue for terpenoids, (2) used GC-MS to identify terpenoids, (3) presented a full profile of compounds in tabular form (i.e., did not exclude unknown compounds; did not only report the most abundant compounds; I did not collect information from original chromatograms), (4) did not juice or burn samples prior to compound extraction, (5) presented data from at least one plant species, and (6) were written in English. To answer questions about climate and phylogeny, I restricted analyses to those papers ($n = 557$ samples, $n = 148$ studies, $n = 206$ species) that also (1) collected plant tissue from naturally occurring organisms (i.e., plants grown in cultivated settings such as greenhouses, farms, and arboretums were excluded), (2) reported exact latitude and longitude coordinates or a searchable municipality to later collect latitude and longitude coordinates, (3) analyzed aboveground tissue samples from inflorescences, foliar tissue such as leaves and needles, fruit, stem, or bark because these samples had some of the highest replication in the database, wider phylogenetic coverage, and are more likely associated to ambient air temperature than belowground tissue, and (4) contained samples from

seeding plants (i.e., gymnosperms and angiosperms). If only one mono- or sesquiterpenoid was detected, and I was calculating Functional Hill Diversity, then the sample size was 533 because it is impossible to calculate with only one compound. A list of studies and plant species included in our analyses can be found in a Zenodo repository under DOI: doi.org/10.5281/zenodo.7757599.

I used R version 4.1.1 and Python version 3.9.12 for all data retrieval and analyses (Python Software Foundation 2021; R Core Team 2021). The following R packages were used for data tidying and wrangling: *dplyr* v1.0.8 (Wickham *et al.* 2022), *googlesheets4* v1.0.0 (Bryan *et al.* 2021), *tidyr* v1.2.0 (Wickham and Girlich 2022), *stringr* v1.4.0 (Wickham 2019), and *tibble* v3.1.6 (Müller and Wickham 2021). These R packages were used for visualization: *tidybayes* v3.0.1 (Kay 2021), *ggplot2* v3.3.5 (Wickham 2016), *tidytrees* v0.3.9 (Yu *et al.* 2022), *ggtree* v3.3.2 (Yu 2020), *ggtreeExtra* v1.4.2 (Xu *et al.* 2021), *ggpubr* v0.4.0 (Kassambara 2020), *scattermore* v0.8 (Kratochvil *et al.* 2022), *gridExtra* v2.3 (Baptiste 2017), and *RColorBrewer* v1.1-3 (Neuwirth 2022). The following R packages were used to calculate credible intervals and model predictions for the visualizations: *performance* v0.9.0 (Lüdtke *et al.* 2021), *emmeans* v1.7.3 (Lenth 2022), and *tidybayes* v3.0.1. The following R packages were used for interfacing R and Python code to collect compound information from the API endpoints: *httr* v1.4.2 (Wickham 2020) and *reticulate* v1.26 (Ushey *et al.* 2022). The following Python packages were used to wrangle and collect data: *gsread* 5.3.2 (Burnashev *et al.* 2022), *pandas* v1.5.2 (McKinney 2010; The pandas Development Team 2022), and *numpy* v1.22.1 (Harris *et al.* 2020).

Following the workflow in Grenié *et al.* (2021), I harmonized plant names with the *lcvp* v3.0.1 and *lcvplants* v2.1.0 R packages (Freiberg *et al.* 2020), which references the Leipzig catalog for vascular plants at a global scale. Plant names that were returned as synonyms were changed to the first result given from the catalog. For those samples that only provided a

municipality name, I used the *tidygeocoder v1.0.5* R package to collect latitude-longitude data using the *geocode* function (Cambon *et al.* 2021). Then, I collected historical climate variables at a 10-minute resolution using WorldClim data (annual mean temperature, *BIO1*, and annual precipitation, *BIO12*, as the average from 1970-2000) at the latitude-longitude coordinates of each sample (Fick and Hijmans 2017).

The name of each compound was standardized in two steps: (1) searching the PubChem application program interface (API) and (2) manual searches across online databases. In Step 1, I ran each compound name through *PubChemPy v1.0.4*, a Python package that retrieves chemical compound information from the United States National Institute of Health PubChem chemical database API, and accepted the first result (Swain *et al.* 2014; Kim *et al.* 2021). In Step 2, for compounds that were not identified from the first result in our Python script, I manually searched the remaining compound names on PubChem's website, Wiley SpectraBase, the Royal Society of Chemistry ChemSpider, and the US National Institute of Standards and Technology (NIST)'s Chemistry WebBook (John Wiley & Sons, Inc. 2022; Linstrom and Mallard 2022; Royal Society of Chemistry 2022).

Once compound names were standardized, I collected the Simplified Molecular Input Line System (SMILES) and/or IUPAC International Chemical Identifier Key (InChIKey) for each compound from each source, as available. These two features detail the structural configuration, chemical components, and other properties based on a series of symbols and letters (Weininger 1988, Heller *et al.* 2015). They were necessary to locate the PubChem fingerprints incorporated in diversity metrics later in the analyses. The SMILES and InChIKey were collected through different mechanisms based on the source. If available on PubChem, I used *PubChemPy* to collect each compound's SMILES and InChIKey. For those compounds

only found on SpectraBase and ChemSpider, I manually copied the values from each compound's HTML webpage between May to July 2022. For compounds found on NIST, I wrote a Python script to pull the SMILES and InChIKey from the NIST URL for each compound. In total, I found the Canonical SMILES or InChIKey for 98.0% of all compound observations across these four sources. Only exact matches for compounds were included in analyses here (e.g., unknown compounds were excluded, even if they were labeled terpenoids).

After collecting these values, each entry was labeled through the *NPClassifier*, which uses deep learning tools to identify the molecular pathway and class structure of a compound (Kim *et al.* 2020), via the *NPCTable* function in the *chemodiv* package. Then, I excluded all chemicals that were not monoterpenoids or sesquiterpenoids. With the *get_compounds* function in *PubChemPy*, I collected information on each compound's substructures through PubChem Substructure Fingerprints v1.3 (National Center for Biotechnology Information 2009). Each fingerprint is a sequence of binary values for the presence or absence of 881 possible substructures in a compound. Fingerprints may not distinguish between compounds that differ in properties such as charge or atom arrangement. Since the full scope of chemical substructure diversity cannot be determined through PubChem fingerprints alone, this approach is a conservative estimate of the full range of mono- and sesquiterpenoid structural diversity produced by plants, and I still present analyses of compound richness and β -diversity to account for differences in these properties. As described below, fingerprints can be treated similarly to species abundance matrices, where the presence or absence of a chemical substructure (e.g., the presence of at least one oxygen molecule) in a compound is analogous to the presence or absence of a species in an ecological community.

Calculating α -diversity and β -diversity

To calculate terpenoid diversity, I considered a sample as the grouping factor (e.g., species, location) presented in each study to develop a list of detected terpenoids. Structural α -diversity was calculated as the Functional Hill Diversity with a diversity order of zero, as has been described in recent work on incorporating structural dissimilarity into compound richness (Petrén *et al.* 2023). Functional Hill Diversity is determined by combining Rao's quadratic entropy, Q , and the equation for Hill Numbers (Rao 1982; Chao *et al.* 2019) to calculate the effective number of the pairwise structural differences between compounds:

$$Q = \sum_{i=1}^S \sum_{j=1}^S d_{ij} p_i p_j$$

$${}^qFD(Q) = \left[\sum_{i,j=1}^S d_{ij} p_i p_j \right]^{1/(1-q)}, \quad q \geq 0, \quad q \neq 1$$

In the above equations, q represents the diversity order. As q increases, structural diversity is increasingly weighted by the proportional abundance of any given compound. d_{ij} represents the compound dissimilarity matrix, which I calculated via the Jaccard distance between any two compounds, calculated with PubChem fingerprints, with the *chemodiv* R package (Figure 3.1; Petré *et al.* 2023). p_i and p_j are the relative abundances of each compound in the sample (S). When $q = 0$, Functional Hill Diversity is equivalent to richness weighted by compound dissimilarity. In all calculations of structural diversity, compounds were treated as being present or absent, and compounds listed as 'trace' in a study were not included. Treating compounds in this qualitative manner reduces the effect of varying extraction and detection methods across studies.

β -diversity was calculated in three ways: (a) compound β -diversity through the Jaccard index for the presence of terpenoid compounds, (b) structural β -diversity through the Jaccard

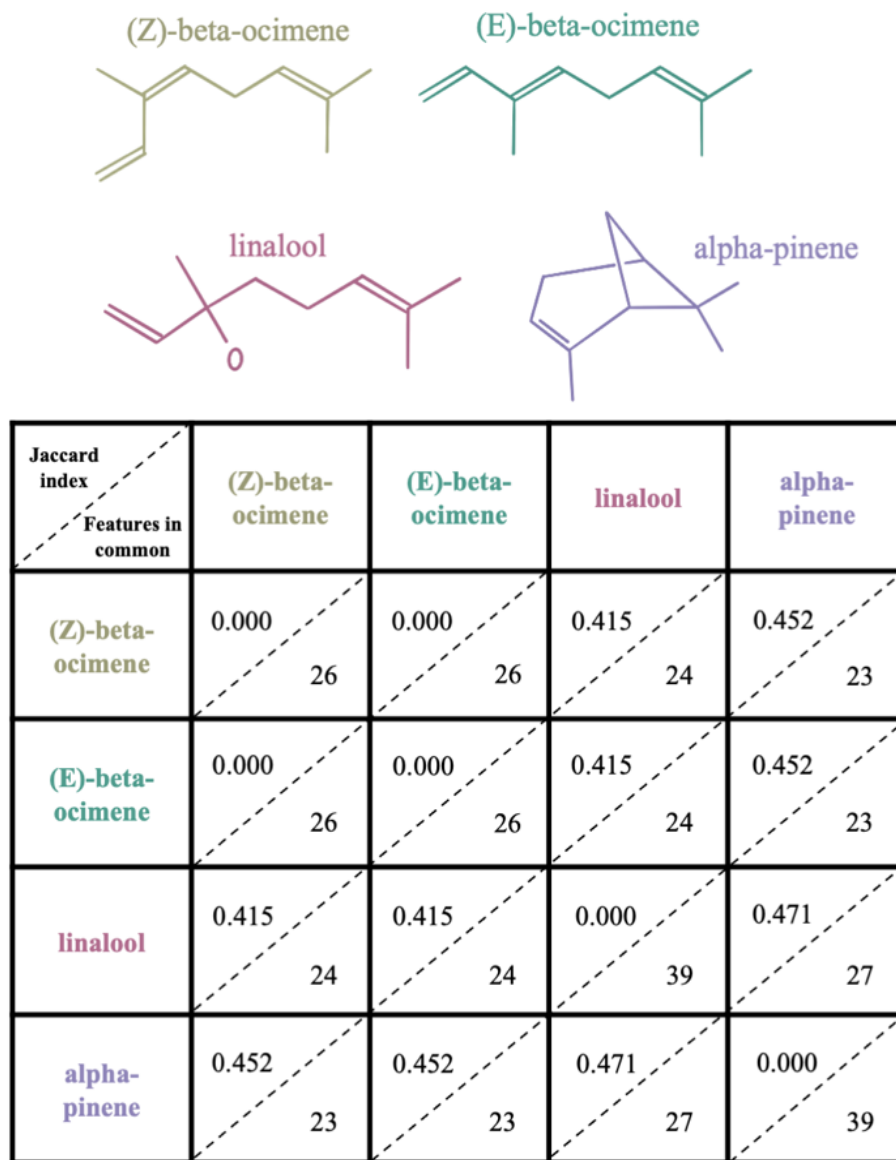


Figure 3.1 Structural dissimilarity and number of structural features in common of four example monoterpene compounds ((Z)- β -ocimene, (E)- β -ocimene, linalool, and α -pinene). Fifty-three substructures in total were identified across the four compounds, and they have 23 substructures in common. Under this fingerprint approach, (E)- β -ocimene and (Z)- β -ocimene are treated identically, and they have two substructures not present in linalool and α -pinene (C=C-C=C and C-C=C-C=C). With a Jaccard index of 0.471, linalool and alpha-pinene are the most different structurally. Linalool has eleven substructures that are not present in any other compound, most notably the hydroxyl group. Finally, α -pinene has twelve unique substructures, principally stemming from the 4-carbon ring.

index for the presence of chemical structures in a sample's terpenoid profile, and (c) structural β -diversity through the Bray-Curtis index for the abundance of chemical structures in a sample's

terpenoid profile. For (b), I determined the presence of any of 881 possible structures as determined by the PubChem fingerprints of each compound. For (c), the Bray-Curtis index was calculated by summing all observations of the 881 possible structures to determine an “abundance” for any given substructure in a sample. Jaccard and Bray-Curtis indices were calculated with the *vegdist* function in the *vegan* v2.6-4 R package (Oksanen *et al.* 2022).

Phylogenetic methods

For phylogenetic analyses, I used GBOTB tree of vascular plants and the *V.PhyloMaker* v0.1.0 R package to calculate phylogenetic distance between the plant species in the sample studies and the phylogenetic variance-covariance matrix (Smith and Brown 2018; Jin and Qian 2019). The phylogenetic variance-covariance matrix was calculated via the *vcv.phylo* function in the *ape* v5.0 R package (Paradis *et al.* 2019). Phylogenetic distance was calculated through the *cophenetic.phylo* function in the *ape* R package.

Statistical methods

Each diversity metric was analyzed using Bayesian regression models through the *brms* package in R using between 1000 and 2500 iterations (i.e., α -diversity models had 2500 iterations; β -diversity models had 1000 iterations), default warmups, and naïve priors (i.e., default values in *brms*; Bürkner 2017). Associations were deemed significant if the 95% credible interval (CI) of the effect size did not overlap zero. Richness was modeled with the count of unique terpenoid compounds and using a Poisson distribution. I modeled structural α -diversity with log-transformed Functional Hill Diversity metric for each sample and using a Gaussian distribution. For model convergence, I centered each of the temperature and precipitation population-level predictors and constrained the values within the same order of magnitude by dividing by 10 and 1000, respectively. The richness and structural α -diversity models had two

group-level effects: one for phylogenetic correlation between species and another for the species itself, which addressed any niche or study-specific effects not accounted by phylogeny, temperature, or precipitation (Bürkner 2022).

For the β -diversity models, I analyzed each pairwise comparison between samples of the same anatomical organ (e.g., leaf samples were compared to leaf samples, floral with floral, fruit with fruit, etc.). I excluded pairwise comparisons within the same species. Three population-level predictors were included in the model: pairwise difference in temperature, pairwise difference in precipitation, and the pairwise phylogenetic distance. For model convergence and efficiency, I centered the temperature, precipitation, and phylogenetic distance predictors and constrained the values by dividing by 10, 1000, and 100 respectively. Both β -diversity metrics (i.e., Jaccard, Bray-Curtis) are constrained between zero and one and thus required running the models under a zero-one inflated beta distribution (compound-level β -diversity) or zero-inflated beta regression distribution (structural β -diversity). I assigned predictors to the parameters that emphasized the explanatory power of values between zero and one. Under the zero-one-inflated beta distribution, the predictors were applied to the mean (μ) and precision (ϕ) parameters, while the zero-one process-based parameters, α and γ , only had intercepts. Similarly, under the zero-inflated beta distribution, I applied the three predictors to the mean (μ) and precision (ϕ) parameters, but not the zero process-based parameter, α .

Results

α -diversity: Richness

In the analyses of terpenoid diversity and climate, temperature was not significantly associated with the sample's terpenoid richness, whereas richness significantly increased by 24.0% (95% CI, 0.4-64.3%) per 100 mm in precipitation (Figure 3.2a-b). Temperature and precipitation

explained about 0.9% (0.0007-0.95%) of the variation in monoterpene and sesquiterpene richness, while the phylogenetic and species group-level effects explained 84.6% (83.6-85.6%) of the variation.

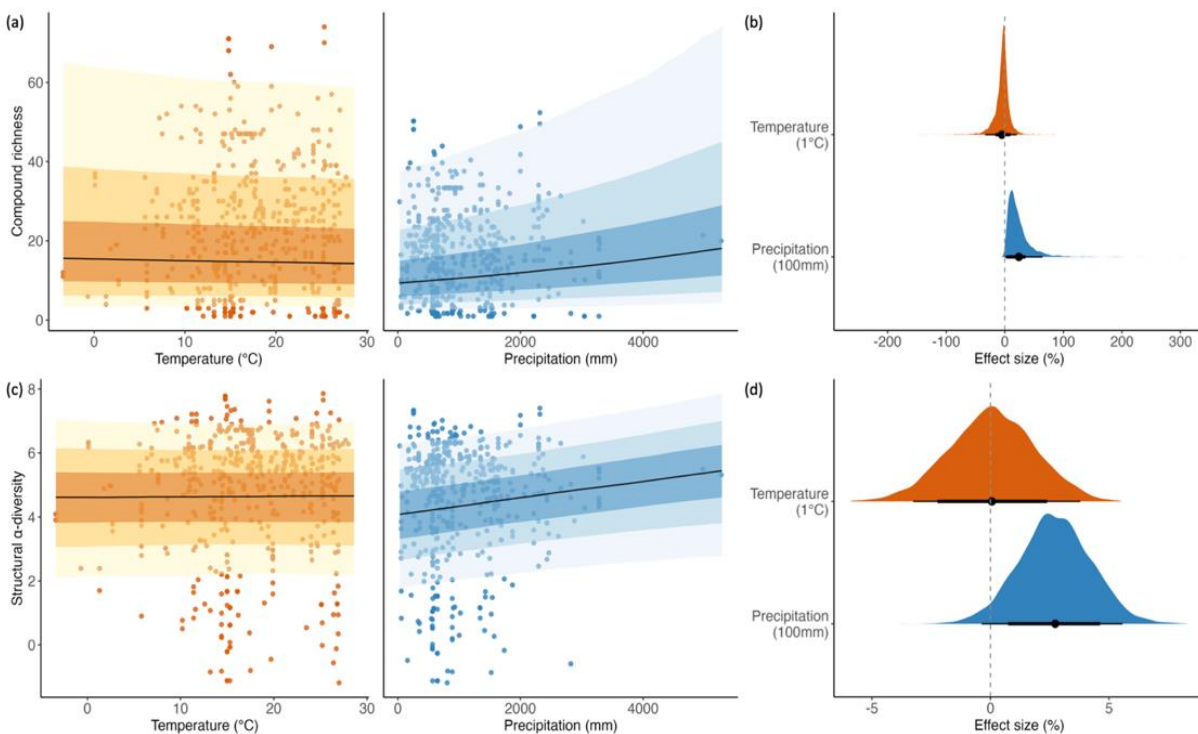


Figure 3.2 α -diversity by climate variables. a) Compound richness across the global temperature and precipitation gradients. b) Effect size posterior densities for compound richness by climate variables. (c) Structural α -diversity (natural log of Functional Hill Diversity with zero diversity order) across the global temperature and precipitation gradients. (d) Effect size posterior densities for structural α -diversity by climate variables. Black trend lines are mean predicted values and the three shaded intervals around represent the 95% (lightest), 80% (medium), and 50% (darkest) credible intervals. Posterior density plots have quartile lines in black underneath representing the 95% (thinnest), 80% (medium), and 50% (thickest) credible intervals. Effect size units are the percentage increase or decrease in the response variable per 1°C or 100mm.

α -diversity: Structural α -diversity

There was no significant association between temperature or precipitation and structural α -diversity (Figure 3.2c-d). When weighing compounds by their structural dissimilarity to calculate Functional Hill Diversity at a zero-diversity order, precipitation and temperature explained 1.3%

(0.002-4.2%) of the structural α -diversity in monoterpenoids and sesquiterpenoids. The phylogenetic and species group-level effects explained 84.1% (82.2-85.7%) of the variation.

β -diversity: Compound presence

After excluding intraspecific comparisons, there were 50,799 pairwise comparisons to test climate and phylogenetic variation with β -diversity. In absolute terms, the effect sizes may appear low. However, Jaccard and Bray-Curtis values are constrained between zero and one, and a 1% increase in these values can represent the difference of a few compounds or substructures that may have significantly different ecological or physiological functions.

A difference of 4°C was significantly associated with a 1.0% (0.9-1.1%) increase in compound-level differences between samples. In other words, if plants were growing in two climates that on average had temperature differences of 4°C, I observed a 1.0% (0.9-1.1%) increase in β -diversity. A difference of 500 mm in precipitation was significantly associated with a 1.0% (0.9-1.1%) increase in compound-level β -diversity. More distantly related species had different monoterpenoids and sesquiterpenoids detected (Figure 3.3a-b). A difference of ten million years in branch length was associated with an 0.03% (0.03-0.04%) increase in β -diversity (Figure 3c-d). Temperature, precipitation, and phylogenetic distance explained 1.7% (1.5-1.8%) of the variation in the compound-level differences between samples (Jaccard index).

β -diversity: Chemical structure presence

A difference of 4°C between samples was significantly associated with a 0.5% (0.4-0.6%) increase in structural presence β -diversity. When incorporating structural information about the compounds, the effect of distance became more negative; a difference of 500 mm in precipitation was marginally, though not statistically, associated with a 0.1% increase (0.005% decrease, 0.2%

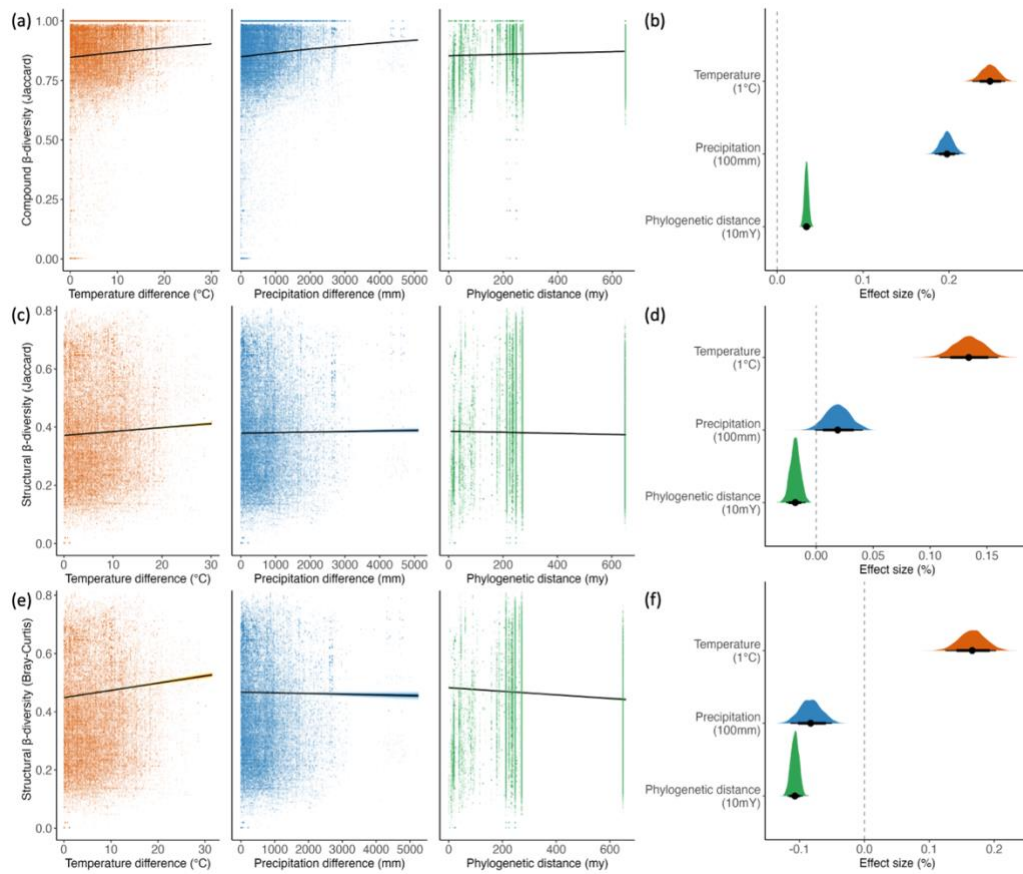


Figure 3.3 Compound and structural β -diversity by climate and phylogenetic variables. (a) Compound β -diversity across gradients of samples' pairwise temperature differences, precipitation differences, and phylogenetic distance. (b) Effect size posterior densities for compound β -diversity by climate and phylogenetic variables. (c) Structural β -diversity based on substructure presence or absence (Jaccard index) across gradients of samples' pairwise temperature differences, precipitation differences, and phylogenetic distance. (d) Effect size posterior densities for structural β -diversity (Jaccard index) by climate and phylogenetic variables. (e) Structural β -diversity based on substructure abundance (Bray-Curtis index) across gradients of samples' pairwise temperature differences, precipitation differences, and phylogenetic distance. (f) Effect size posterior densities for structural β -diversity (Bray-Curtis index) by climate and phylogenetic variables. Black trend lines are mean predicted values and the three shaded intervals around represent the 95% (lightest), 80% (medium), and 50% (darkest) credible intervals. Posterior density plots have quartile lines in black underneath representing the 95% (thinnest), 80% (medium), and 50% (thickest) credible intervals. Effect size units are the percentage increase or decrease in the response variable per 1°C or 100mm.

increase) in structural β -diversity. Unlike compound β -diversity, more closely related species had different structures present in their monoterpenoid and sesquiterpenoid profiles. A 0.20% (0.1-0.3%) increase in the structural similarity corresponded with every ten-million-year increase in phylogenetic distance (Figure 3.3c-d). Temperature, precipitation, and phylogenetic distance explained 0.2% (0.1-0.3%) of the total variation in structural β -diversity (Jaccard index).

β -diversity: Chemical structure abundance

The relationship between temperature and β -diversity was strongest when accounting for structural abundance; a difference of 4°C between samples was associated with a 0.7% (0.5-0.8%) increase in structural β -diversity. Unlike the result in compound β -diversity, samples growing in environments with more different precipitation levels showed more similar abundances in substructures. A difference of 500 mm of precipitation in the environments between two samples was associated with a 0.4% (0.3-0.6%) decrease in structural β -diversity. Phylogenetic distance had a significant association with this diversity metric, with a 0.1% (0.1-0.1%) increase in structural β -diversity associated with a ten-million-year increase in the branch length between samples (Figure 3.3e-f). When accounting for the abundance (counts) of these structures, precipitation, temperature, and phylogenetic distance explained 0.7% (0.5-0.8%) of the variation in structural abundance β -diversity (Bray-Curtis index).

Discussion

This study presents the first geographically extensive meta-analysis that evaluates how climate and phylogenetic distance affect the diversity and structure of a specific set of phytochemical compounds. This study is also one of the first studies to tie together various sources of information on compound presence and structures to understand how plant chemical

mixtures may differ in ways not immediately apparent from the compound names alone. For example, by incorporating structural information on the terpenoid profiles into the β -diversity metrics, I observed opposite patterns for precipitation and phylogenetic distance between compound presence and substructure abundance. Limiting the comparisons only to those samples that followed similar methodological protocols (e.g., GC-MS identification of terpenoids) that were relevant to the ecological function and physicochemical constraints of their release into the environment (e.g., GC-MS identification of volatile and semi-volatile compounds) was also a successful approach. Diving into possible reasons driving these associations, or non-associations, will further highlight the significance of the work.

Since monoterpenoids and sesquiterpenoids are often only functionally useful as volatile and semi-volatile compounds, the lack of association between temperature and α -diversity is surprising. As indicated by the results across all measures of β -diversity, the specific compounds and substructures significantly differ across a gradient of temperature differences. Together, the results from α - and β -diversity analyses suggest the identities of the compounds and substructures are associated more strongly with temperature than diversity *per se*. Warmer temperatures may not allow for greater diversity if a volatile set of compounds degrades more quickly than at lower temperatures, reducing their role in ecological interactions. While warmer temperatures may increase terpenoid production and emissions, temperatures above thermal optima can shut down key enzymatic reactions (Loreto *et al.* 2006). Finally, there is evidence that some isoprenoids (e.g., terpenoids), may aid thermal stress tolerance (Loreto *et al.* 1998). If these compounds, such as isoprene and some monoterpenoids, help protect plants' cellular membranes when facing thermal stress, then this adaptive role may be stronger than selecting compounds that contribute to a more diverse chemical bouquet.

My finding of greater terpenoid α -diversity at higher precipitation levels should prompt further investigation into the role of stomatal closure and presence of highly soluble oxygenated terpenoids. My analyses did not separate oxygenated and non-oxygenated compounds, despite oxygenated compounds having higher solubility in plant tissue and being less likely to diffuse into the external environment without stomatal opening (Niinemets et al. 2004). Wetter environments, such as those found at lower latitudes, are often hypothesized to have greater herbivory pressure, which may represent an indirect effect behind these patterns. The β -diversity patterns were variable at the compound and substructure levels; while compound β -diversity was lower between samples of more similar precipitation levels, structural β -diversity decreased with more similar precipitation levels. Overall, plants growing in more similar precipitation environments may have the same baseline compounds that function given the physiological constraints for those conditions. However, under extreme precipitation conditions, plants may have similar substructures that either help them tolerate those stressors or are a byproduct of the stress caused by both drought and wet conditions.

Phylogenetic distance between samples was positively associated with compound β -diversity. At least in monoterpenoids and sesquiterpenoids, plants that are more closely related tend to have more similar terpenoid bouquets. However, by incorporating the substructural composition of these compounds into the structural β -diversity calculations, a more nuanced story arises. Given the negative association between phylogenetic distance and structural β -diversity, more related plant species *may* produce a few substructures that distinguish themselves from their relatives. The specificity in reception to these compounds in ecological interactions and the large array of phytochemicals present in a plant community may necessitate these substructural differences in a community of closely related plants. Nevertheless, from these

analyses, I cannot draw conclusions about whether these compounds and substructures are more diverse in some species groups than others. Future analyses with these data may identify the origins of certain common monoterpenoids and sesquiterpenoids as well as attempt to understand if terpenoid diversity begets species diversity or *vice versa*.

The application of ecological diversity metrics to phytochemical compounds and their constituent structures represents a novel avenue to understand how diversification arose and how compound bouquets influence plant, insect, and microbial interactions. I focused on diversity *per se* as it relates to temperature, precipitation, and phylogenetic relatedness. Rather than using a quantitative measure of relative compound abundance, diversity was calculated through compound presence. When accounting for compound concentration (i.e., the abundance of a compound), these relationships may differ because some compounds, such as α -pinene, can dominate the relative amounts in a terpenoid profile. More complex models incorporating the methodological differences in terpenoid detection (e.g., compound polarity, extraction method, machine run time) will be required to increase the order q with these data.

Investigating monoterpenoid and sesquiterpenoid diversity as opposed to other chemical classes was intentional. These compounds are produced within connected biosynthetic pathways. By focusing analyses on these groups, possible explanations for observed patterns are more specific than if we analyzed the diversity of phytochemicals that are encoded across different segments of genomes, produced via different pathways, and/or require different resources. Monoterpenoids and sesquiterpenoids are more uniformly detectable and widely described than heavier weight terpenoids like diterpenoids and triterpenoids. They are also more easily volatilized and are composed of different substructures than other phytochemicals. Therefore, their diversity may follow different geographical, climatic, and phylogenetic relationships than

those found with other chemical classes, like tannins, flavonoids, or phenolics (Pearse and Hipp 2012, Moreira *et al.* 2018, Agrawal *et al.* 2022). If other chemical compound classes or subclasses covary with phylogenetic distance and environmental gradients, attention should be given to contextualize how those compounds are synthesized and encoded in the plant genome before extrapolating broader conclusions.

Plants have adapted a variety of mechanisms to protect themselves, communicate, and interact with other organisms, including the production of secondary metabolites like monoterpenoids and sesquiterpenoids. The identity and synthesis of these chemicals has come into clearer view (Pickersky and Raguso 2018), but the “where,” “when,” and “why” these chemical are so diverse remain largely unresolved. By analyzing such a broad range of abiotic conditions and phylogenetic distances, I successfully identified *where* along climate and phylogenetic distance gradients monoterpenoid and sesquiterpenoid α -diversity and β -diversity is greatest. These results lead to many questions about *why*, such as (1) does the ambient air temperature directly constrain or promote plant monoterpenoids and sesquiterpenoids diversity, or is there some other cofactor like animal diversity or metabolic activity at warmer temperatures driving this pattern? (2) do plants that are more closely related phylogenetically have more similar antagonist and mutualist communities that drive the patterns we see here? If available, future environmental analyses could include mean atmospheric pressure as a predictor of structural diversity, as it is a complementary determinant for volatility and the diffusion gradient between internal plant tissue and the external environment. These analyses could also account for daily and annual variation in environmental predictors. We could also analyze the occurrences of these specific substructures (e.g., those substructures that lead to higher solubility) to determine the physicochemical and ecological correlates of terpenoid diversity beyond these broad metrics.

Experimental approaches coupled with the prior knowledge of these broad-scale patterns in phytochemical diversity will lead us probably to more answers and certainly more questions. But for now, we know that hypotheses should address the biotic *and* abiotic environment to understand universal patterns of phytochemical diversity in nature.

REFERENCES

- Abdala-Roberts L, Moreira X, Rasmann S, et al (2016) Test of biotic and abiotic correlates of latitudinal variation in defences in the perennial herb *Ruellia nudiflora*. *Journal of Ecology* 104:580–590. <https://doi.org/10.1111/1365-2745.12512>
- Adler LS (2000) The ecological significance of toxic nectar. *Oikos* 91:409–420. <https://doi.org/10.1034/j.1600-0706.2000.910301.x>
- Alessio GA, Peñuelas J, Llusà J, et al (2008) Influence of water and terpenes on flammability in some dominant Mediterranean species. *Int J Wildland Fire* 17:274. <https://doi.org/10.1071/WF07038>
- Agrawal AA, Espinosa del Alba L, López-Goldar X, et al (2022) Functional evidence supports adaptive plant chemical defense along a geographical cline. *Proceedings of the National Academy of Sciences* 119:e2205073119. <https://doi.org/10.1073/pnas.2205073119>
- Baptiste A (2017) gridExtra: Miscellaneous functions for “grid” graphics. <https://cran.r-project.org/web/packages/gridExtra/index.html>
- Becerra JX (1997) Insects on plants: Macroevolutionary chemical trends in host use. *Science* 276:253–256. <https://doi.org/10.1126/science.276.5310.253>
- Berenbaum M, Feeny P (1981) Toxicity of angular furanocoumarins to swallowtail butterflies: escalation in a coevolutionary arms race? *Science* 212:927–929. <https://doi.org/10.1126/science.212.4497.927>
- Bleeker PM, Diergaarde PJ, Ament K, et al (2009) The role of specific tomato volatiles in tomato-whitefly interaction. *Plant Physiology* 151:925–935. <https://doi.org/10.1104/pp.109.142661>
- Bryan J (2021) googlesheets4: Access Google Sheets using the Sheets API V4. <https://googlesheets4.tidyverse.org>
- Bürkner P (2017) brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80:1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner P (2022) Estimating Phylogenetic Multilevel Models with brms. https://cran.r-project.org/web/packages/brms/vignettes/brms_phylogenetics.html. Accessed 20 Dec 2022
- Burnashev A (2022) gspread. <https://github.com/burnash/gspread/>
- Cambon J, Hernangómez D, Belanger C, Possenriede D (2021) tidygeocoder: An R package for geocoding. *Journal of Open Source Software* 6:3544. <https://doi.org/10.21105/joss.03544>

- Chao A, Chiu C-H, Villéger S, et al (2019) An attribute-diversity approach to functional diversity, functional beta diversity, and related (dis)similarity measures. *Ecological Monographs* 89:e01343. <https://doi.org/10.1002/ecm.1343>
- Chen C, Song Q (2008) Responses of the pollinating wasp *Ceratosolen solmsi marchali* to odor variation between two floral stages of *Ficus hispida*. *J Chem Ecol* 34:1536–1544. <https://doi.org/10.1007/s10886-008-9558-4>
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* 66:212–229. <https://doi.org/10.1111/j.1365-3113.2011.04520.x>
- Christianson DW (2008) Unearthing the roots of the terpenome. *Current Opinion in Chemical Biology* 12:141–150. <https://doi.org/10.1016/j.cbpa.2007.12.008>
- Defosse E, Pitteloud C, Descombes P, et al (2021) Spatial and evolutionary predictability of phytochemical diversity. *Proceedings of the National Academy of Sciences* 118:e2013344118. <https://doi.org/10.1073/pnas.2013344118>
- Després L, David J-P, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution* 22:298–307. <https://doi.org/10.1016/j.tree.2007.02.010>
- Dyer LA (2018) Multidimensional diversity associated with plants: a view from a plant–insect interaction ecologist. *American Journal of Botany* 105:1439–1442. <https://doi.org/10.1002/ajb2.1147>
- Ehrlich PR, Raven PH (1964) Butterflies and plants: A study in coevolution. *Evolution* 18:586–608. <https://doi.org/10.1111/j.1558-5646.1964.tb01674.x>
- Fall R, Monson RK (1992) Isoprene emission rate and intercellular isoprene concentration as influenced by stomatal distribution and conductance 1. *Plant Physiol* 100:987–992
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37:4302–4315. <https://doi.org/10.1002/joc.5086>
- Freiberg M, Winter M, Gentile A, et al (2020) LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Scientific Data* 7:416. <https://doi.org/10.1038/s41597-020-00702-z>
- Gershenzon J, Dudareva N (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology* 3:408–414. <https://doi.org/10.1038/nchembio.2007.5>
- Glassmire AE, Philbin C, Richards LA, et al (2019) Proximity to canopy mediates changes in the defensive chemistry and herbivore loads of an understory tropical shrub, *Piper kelleyi*. *Ecology Letters* 22:332–341. <https://doi.org/10.1111/ele.13194>

- Grenié M, Berti E, Carvajal-Quintero J, et al (2023) Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. *Methods in Ecology and Evolution* 14:12-25 <https://doi.org/10.1111/2041-210X.13802>
- Harris CR, Millman KJ, van der Walt SJ, et al (2020) Array programming with NumPy. *Nature* 585:357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Heller SR, McNaught A, Pletnev I, et al (2015) InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics* 7:23. <https://doi.org/10.1186/s13321-015-0068-4>
- Jin Y, Qian H (2019) V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. *Ecography* 42:1353–1359. <https://doi.org/10.1111/ecog.04434>
- John Wiley & Sons, Inc. (2022) SpectraBase. <https://spectrabase.com/>. Accessed 2022
- Junker RR, Kuppler J, Amo L, et al (2018) Covariation and phenotypic integration in chemical communication displays: biosynthetic constraints and eco-evolutionary implications. *New Phytol* 220:739–749. <https://doi.org/10.1111/nph.14505>
- Kassambara A (2020) ggpubr: “ggplot2” Based Publication Ready Plots. <https://github.com/kassambara/ggpubr>
- Kay M (2021) tidybayes: Tidy data and geoms for Bayesian models. <http://mjskay.github.io/tidybayes/articles/tidybayes.html>
- Kessler A, Kalske A (2018) Plant secondary metabolite diversity and species interactions. *Annu Rev Ecol Evol Syst* 49:115–138. <https://doi.org/10.1146/annurev-ecolsys-110617-062406>
- Kigathi RN, Weisser WW, Reichelt M, et al (2019) Plant volatile emission depends on the species composition of the neighboring plant community. *BMC Plant Biol* 19:58. <https://doi.org/10.1186/s12870-018-1541-9>
- Kim H, Wang M, Leber C, et al (2020) NPClassifier: A deep neural network-based structural classification tool for natural products. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.12885494.v1>
- Kim S, Chen J, Cheng T, et al (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* 49:D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>
- Kratochvil M (2022) scattermore: Scatterplots with more points. <https://github.com/exaexa/scattermore>
- Lenth R (2022) emmeans: Estimated marginal means, aka least-squares means. <https://github.com/rvlenth/emmeans>

- Linstrom P, Mallard W (2022) NIST Chemistry WebBook, NIST Standard Reference Database Number 69. National Institute of Standards and Technology, Gaithersburg, Maryland
- Llusà J, Peñuelas J (2000) Seasonal patterns of terpene content and emission from seven Mediterranean woody species in field conditions. *American Journal of Botany* 87:133–140. <https://doi.org/10.2307/2656691>
- Llusà J, Peñuelas J, Alessio GA, Estiarte M (2006) Seasonal contrasting changes of foliar concentrations of terpenes and other volatile organic compound in four dominant species of a Mediterranean shrubland submitted to a field experimental drought and warming. *Physiologia Plantarum* 127:632–649. <https://doi.org/10.1111/j.1399-3054.2006.00693.x>
- Llusà J, Peñuelas J, Sardans J, et al (2010) Measurement of volatile terpene emissions in 70 dominant vascular plant species in Hawaii: aliens emit more than natives. *Global Ecology and Biogeography* 19:863–874. <https://doi.org/10.1111/j.1466-8238.2010.00557.x>
- Loreto F, Schnitzler J-P (2010) Abiotic stresses and induced BVOCs. *Trends in Plant Science* 15:154–166. <https://doi.org/10.1016/j.tplants.2009.12.006>
- Lüdecke D, Ben-Shachar M, Patil I, et al (2021) performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software* 6:3139
- McKinney W (2010) Data Structures for Statistical Computing in Python. Austin, Texas, pp 56–61
- Moreira X, Abdala-Roberts L, Parra-Tabla V, Mooney KA (2015) Latitudinal variation in herbivory: influences of climatic drivers, herbivore identity and natural enemies. *Oikos* 124:1444–1452. <https://doi.org/10.1111/oik.02040>
- Moreira X, Castagneyrol B, Abdala-Roberts L, et al (2018) Latitudinal variation in plant chemical defences drives latitudinal patterns of leaf herbivory. *Ecography* 41:1124–1134. <https://doi.org/10.1111/ecog.03326>
- Müller K, Wickham H (2021) tibble: Simple data frames. <https://tibble.tidyverse.org/>
- National Center for Biotechnology Information (2009) PubChem subgraph fingerprint. <https://pubchem.ncbi.nlm.nih.gov/>. Accessed 2022-2023.
- Neuwirth E (2022) RColorBrewer: ColorBrewer palettes. <https://cran.r-project.org/web/packages/RColorBrewer/index.html>
- Niinemets Ü, Reichstein M (2003) Controls on the emission of plant volatiles through stomata: Differential sensitivity of emission rates to stomatal closure explained. *Journal of Geophysical Research: Atmospheres* 108. <https://doi.org/10.1029/2002JD002620>
- Niinemets Ü, Loreto F, Reichstein M (2004) Physiological and physicochemical controls on foliar volatile organic compound emissions. *Trends in Plant Science* 9:180–186. <https://doi.org/10.1016/j.tplants.2004.02.006>

- Oksanen J, Simpson G, Blanchet FG, et al (2022) vegan: Community ecology package. <https://github.com/vegandevs/vegan>
- Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Pearse IS, Hipp AL (2012) Global Patterns of Leaf Defenses in Oak Species. *Evolution* 66:2272–2286. <https://doi.org/10.1111/j.1558-5646.2012.01591.x>
- Petrén H, Köllner TG, Junker RR (2023) Quantifying chemodiversity considering biochemical and structural properties of compounds with the R package chemodiv. *New Phytologist* 237:2478–2492. <https://doi.org/10.1111/nph.18685>
- Philbin CS, Dyer LA, Jeffrey CS, et al (2022) Structural and compositional dimensions of phytochemical diversity in the genus *Piper* reflect distinct ecological modes of action. *Journal of Ecology* 110:57–67. <https://doi.org/10.1111/1365-2745.13691>
- Pichersky E, Raguso RA (2018) Why do plants produce so many terpenoid compounds? *New Phytol* 220:692–702. <https://doi.org/10.1111/nph.14178>
- Python Software Foundation (2021) Python. <https://www.python.org/>
- R Core Team (2021) R: A language and environment for statistical computing. <https://www.r-project.org/>
- Rao CR (1982) Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology* 21:24–43. [https://doi.org/10.1016/0040-5809\(82\)90004-1](https://doi.org/10.1016/0040-5809(82)90004-1)
- Richards LA, Dyer LA, Forister ML, et al (2015) Phytochemical diversity drives plant–insect community diversity. *Proceedings of the National Academy of Sciences* 112:10973–10978. <https://doi.org/10.1073/pnas.1504977112>
- Royal Society of Chemistry (2022) ChemSpider. <http://www.chemspider.com/>. Accessed 2022
- Salazar D, Lokvam J, Mesones I, et al (2018) Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat Ecol Evol* 2:983–990. <https://doi.org/10.1038/s41559-018-0552-0>
- Salazar D, Marquis RJ (2022) Testing the role of local plant chemical diversity on plant–herbivore interactions and plant species coexistence. *Ecology* 103:e3765. <https://doi.org/10.1002/ecy.3765>
- Schwab W, Fuchs C, Huang F (2013) Transformation of terpenes into fine chemicals. *Eur J Lipid Sci Technol* 115:3–8. <https://doi.org/10.1002/ejlt.201200157>
- Swain M, Kurniawan E, Powers Z, et al (2014) PubChemPy. <https://pubchempy.readthedocs.io/en/latest/>

- Tetali SD (2019) Terpenes and isoprenoids: a wealth of compounds for global use. *Planta* 249:1– 8. <https://doi.org/10.1007/s00425-018-3056-x>
- The Pandas Development Team (2022) pandas-dev/pandas: Pandas. <https://pandas.pydata.org/>
- Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology* 9:297–304. <https://doi.org/10.1016/j.pbi.2006.03.014>
- Tholl D (2015) Biosynthesis and Biological Functions of Terpenoids in Plants. In: Schrader J, Bohlmann J (eds) *Biotechnology of Isoprenoids*. Springer International Publishing, Cham, pp 63–106
- Ushey K, Allaire J, Tang Y (2022) reticulate: Interface to “Python”. <https://rstudio.github.io/reticulate/>
- Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model* 28:31–36. <https://doi.org/10.1021/ci00057a005>
- Went FW (1960) Blue hazes in the atmosphere. *Nature* 187:641–643. <https://doi.org/10.1038/187641a0>
- Whitehead SR, Bass E, Corrigan A, et al (2021) Interaction diversity explains the maintenance of phytochemical diversity. *Ecology Letters* ele.13736. <https://doi.org/10.1111/ele.13736>
- Wickham H (2016) *Elegant Graphics for Data Analysis*. Springer-Verlag New York
- Wickham H (2019) stringr: Simple, consistent wrappers for common string operations. <https://stringr.tidyverse.org/>
- Wickham H (2020) httr: Tools for working with URLs and HTTP. <https://httr.r-lib.org/>
- Wickham H, François R, Henry L, Müller K (2022) dplyr: A grammar of data manipulation. <https://dplyr.tidyverse.org/>
- Wickham H, Girlich M (2022) tidyr: Tidy messy data. <https://tidyr.tidyverse.org/>
- Xu S, Dai Z, Guo P, et al (2021) ggtreeExtra: Compact visualization of richly annotated phylogenetic data. 38:4039–4042. *Molecular Biology and Evolution*. <https://doi.org/10.1093/molbev/msab166>
- Yu G (2020) Using ggtree to visualize data on tree-like structures. *Current Protocols in Bioinformatics* 69. <https://doi.org/10.1002/cpbi.96>
- Yu G (2022) tidytree: A tidy tool for phylogenetic tree data manipulation. <https://github.com/YuLab-SMU/tidytree>

CHAPTER 4:

TERPR V1.0.0: A DATABASE OF PLANT MONOTERPENOIDS AND SESQUITERPENOIDS

Abstract

Monoterpenoids and sesquiterpenoids are ubiquitous phytochemicals that serve as warning signals of stress, attractants to pollinators, and natural products for many human uses. Although thousands of these compounds have been identified, their diversity and distribution within and across species. have not been centralized in a database due to methodological variation in their identification and quantification, differences in compound naming conventions, and the unstructured nature of data in published studies. Here I present *terpr v1.0.0*, the first iteration of a global database of monoterpenoids and sesquiterpenoids covering 1178 plant species across 5107 samples from 1227 studies. The relational database consists of 86 features across six tables that describe the detected monoterpenoids and sesquiterpenoids, plant growth and tissue collection methods, and analytical chemistry analyses in each study. The database can be used to identify broad patterns in the diversity and distribution of these compounds across plants, to answer long-standing questions in ecology and evolution, as well as to model how a database of complex, unstructured information can be engineered using modern technological and coding tools.

Background & Summary

The centralization of scientific literature through publishing websites, online databases, and search engines has unlocked vast potential for the synthesis of existing knowledge. The identification and publication of biologically synthesized chemical compounds has revolutionized pharmaceutical discovery, the study of ecology, and our understanding of biogenic effects on climate (Hartmann 2007, Tetali 2019; Zu *et al.* 2020; Weber *et al.* 2022). For millions of years, plants have been incubators of chemical diversity—producing thousands of unique compounds that ward off insect attack, prevent the establishment of competitors, and signal times of stress to neighbors (Becerra 1997; Junker *et al.* 2018; Kalske *et al.* 2019). Those compounds that are not responsible for plant primary metabolism, known as secondary metabolites, have exhibited particularly high diversification in form and function over the course of plant evolution (Kessler and Kalske 2018). Despite the plethora of studies describing phytochemical profiles across thousands of plant species, the unstructured (i.e., not in a form that can be readily analyzed) nature of these data has prevented synthetic approaches to identify broader patterns and the discovery of novel metabolites that can explain ecological interactions and benefit humans. Compiling studies identifying two biochemical subclasses—monoterpenoids and sesquiterpenoids—across many plant species, I present a central source, *terpr v1.0.0*, to reference, analyze, and serve as a model for the curation of large quantities of phytochemical knowledge.

Monoterpenoids and sesquiterpenoids are ideal chemical superclasses of plant chemicals with which to break down the barriers obscuring high level patterns in phytochemical diversity and distribution. These compounds are ubiquitous in plants and their biosynthesis shares similar pathways across species (Chen *et al.* 2011). However, terpenoids can differ from each other in

configuration, polarity, and the addition of functional groups. They are also easily and uniformly detected in a well-established and accepted laboratory method—gas chromatography-mass spectrometry (GC-MS). Prior to qualitative identification with GC-MS, compounds are extracted from tissues through methods such as hydrodistillation, extraction via organic solvent, and static and dynamic headspace sampling. GC-MS is often paired with gas chromatography-flame ionization detection (GC-FID) or some dynamic sampling method and direct injection to reliably quantify the concentration of a given compound in a plant’s chemical profile. Alternatively, headspace solid-phase microextraction (HS-SPME), a type of static headspace sampling method, has been shown to give more variable concentrations than the other methods (Tholl et al. 2006). Given their light molecular weight relative to larger terpenoids and other plant-produced hydrocarbons, monoterpenoids and sesquiterpenoids are routinely detected across the most common GC run times. These and other methodological differences may present a significant source of variation between studies. Given the numerous other methods for detecting phytochemicals, such as liquid chromatography-mass spectrometry and newly expanding metabolomics techniques, focusing on this single identification approach allows for synthetic comparisons among the numerous studies using GC-MS.

Existing compendia of phytochemicals typically present data as “natural products” pertaining to human use for food, therapeutics, and/or pharmaceuticals. A recent review of natural product databases found 92 open-access resources made available in the last ~20 years (Sorokina and Steinbeck 2020). These databases include, but are not limited to, publicly available online repositories such as Dr. Duke’s Phytochemical and Ethnobotanical Database (Duke and Bogenschutz 1994) and Natural Products ALERT Database (Loub et al. 1985), and newer, geographically specific repositories published as data papers, such as the database of

Indian Medicinal Plants, Phytochemistry, and Therapeutics (Mohanraj et al. 2018). As the natural products industry boomed with information, the nascent field of biocuration also launched in the last decade (Howe et al. 2008; International Society for Biocuration 2018). Primarily focused on the collection, maintenance, and operability of large amounts of biomedical data and gene ontologies, biocuration presents an interesting parallel to the work of phytochemical data curation as it emphasizes the effectiveness of amalgamations across many knowledge bases.

Based on the standards set by this foundational work, I present a relational database of plant monoterpenoids and sesquiterpenoids identified via GC-MS, *terpr v1.0.0*, for use in our basic understanding of phytochemical diversity and for natural product discovery. The *terpr v1.0.0* database consists of the highest resolution database for the collection of plant tissue, compound detection methodology, and the locations of two specific chemical subclasses known to date. With interoperability across other chemical databases, such as PubChem (Kim et al. 2021) and ChemSpider (Royal Society of Chemistry 2022), this database serves in two capacities: (1) as a source of central knowledge to test hypotheses about the ecology, evolution, and function of plant monoterpenoids and sesquiterpenoids and (2) a model for the curation of databases of other subclasses of phytochemicals following FAIR principles so that data is findable, accessible, interoperable, and reproducible (Sansone et al. 2019; Jacobsen et al. 2020).

Methods

Study inclusion

Data were collected from primary literature describing the chemical profiles of plants using GC-MS and filtered through a series of criteria (Figure 4.1). I searched ISI Web of Science for the broadest search of papers using a combination of terms indicating GC-MS usage and the evaluation of terpenes and terpenoids (Table 4.1). In October 2020, I organized a working group

Search terms	Result count	Date searched
terpene* gc*ms	60	23-Nov-2020
terpene* gc-ms	2121	23-Nov-2020
terpenoid* gc*ms	36	23-Nov-2020
terpenoid* gc-ms	1472	23-Nov-2020
terpene* ecolog*	928	01-Dec-2020
terpene* herbivor*	666	01-Dec-2020
terpenoid* ecolog*	822	01-Dec-2020
terpenoid* herbivor*	582	01-Dec-2020

Table 4.1 Details on ISI Web of Science search.

of four graduate students and four faculty across five universities who had specialized research interests in terpenoids, chemical ecology, and quantitative methods. This working group helped determine the study inclusion criteria and the features collected from each paper. Since there

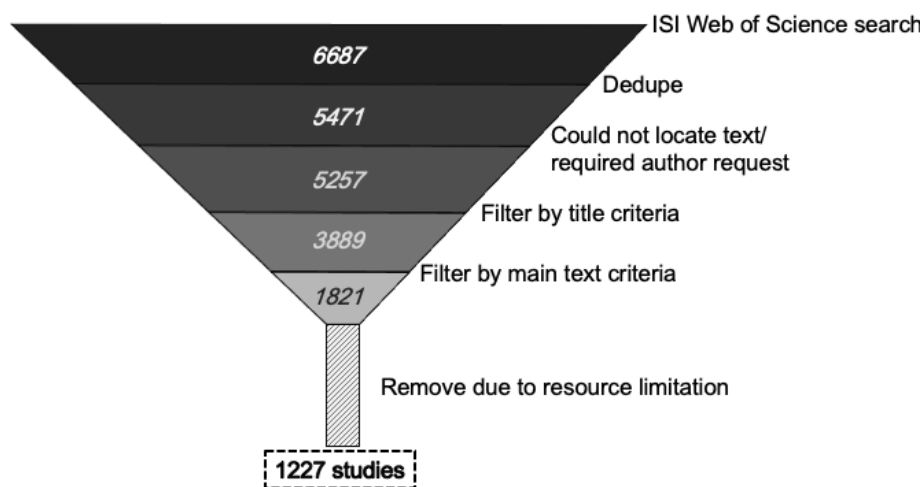


Figure 4.1 Filter with the study count at each stage of criteria evaluation.

were specific interests in the working group producing this database on the ecological and evolutionary implications of these lightweight terpenoids, I also searched terms for ecology and herbivory/herbivores.

Each paper underwent an initial review of the title for exclusion if the paper mentioned only animal, fungus, bacteria, animal product (i.e., milk), detritus, fermented material (e.g., wine or beer) as the study focus. For studies that were not initially excluded or that were ambiguously titled under this criterion, a more detailed review followed. In this second review, I only included papers that (a) detected chemical compounds via GC-MS, (b) presented detected compounds in tabular form, (d) did not exclude unknown compounds (i.e., peak was detected but not formally identified), (e) grouped samples only within a single plant species, (f) did not juice or burn samples, (g) collected initial tissue samples from living plants, (h) did not indicate only the most abundant compounds were reported (i.e., top 10 most abundant compounds by emissions rate), and (i) were written in English. Dried samples were included. If a paper presented some plants that were grown in culture or *in vitro*, those samples were excluded from the database. Finally, as data collection progressed, 584 papers were not included due to resource and time constraints because those papers were either (a) written with ambiguous details about tissue sampling and/or chemical methodology or (b) required labor-intensive manual data collection of the detected compounds and/or methodology. In future database iterations, those papers that require manual data collection (i.e., table in PDF was not easily copied) could be run through natural language processing pipelines or manually transcribed if more labor resources are available.

Data collection

Each table in the relational database required separate collection protocols, with custom tools necessary for the *Concentration* and *Chemical Methods* tables. The generalized data collection workflow (Figure 4.2) for an article that met my inclusion criteria was: (a) collect sample information from the main text and/or supplemental tables for each column in the article's relevant table(s); (b) collect treatment, location, and chemical methods (in no particular order) from the article's main text and/or supplemental tables while collecting, transforming, and standardizing concentration values and compound names; (c) organize the studies' samples by

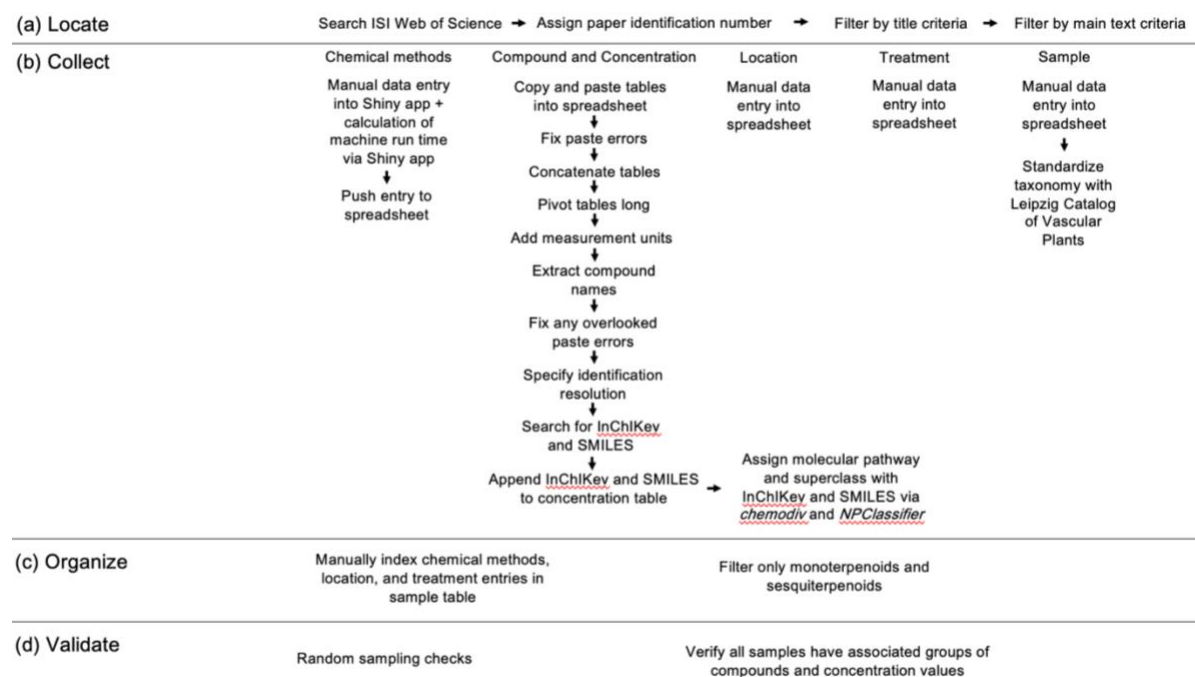


Figure 4.2 Workflow for each study, from filtering by inclusion criteria to validation.

Connecting the *sample_id* as a secondary key to the primary keys for the *Chemical Methods*, *Location*, and *Treatment* tables; and (d) random sample checks and verification that every group of concentration values (i.e., sample) has an associated *sample_id*.

I present each table's protocol in alphabetical order, separated by relational database table. Each *terpr* table can be related through primary and secondary keys (e.g., "sample_id",

“location_id”, “paper_id”). Here, I review some key features and their construction, but full lists of every table feature, their definition, and possible values are in Tables 2-7.

R version 4.1.1 and Python version 3.9.12 were used for data retrieval and analyses (Python Software Foundation 2021; R Core Team 2021). Data were cleaned and organized with the following packages: *dplyr* v1.0.8 (Wickham *et al.* 2022), *googlesheets4* v1.0.0 (Bryan *et al.* 2021), *stringr* v1.4.0 (Wickham 2019), *tibble* v3.1.6 (Müller and Wickham 2021), and *tidyr* v1.2.0 (Wickham and Girlich 2022). The following R packages visualized the summary data presented in the database: *ggplot2* v3.3.5 (Wickham 2016), *ggrepel* v0.9.1 (Slowikowski 2021), *scattermore* v0.8 (Kratochvil *et al.* 2022) and *viridis* v0.9.1 (Garnier *et al.* 2021). The following R packages helped connect R and Python code and/or collect compound information from the API software development kits: *httr* v1.4.2 (Wickham 2020) and *reticulate* v1.26 (Ushey *et al.* 2022). The following Python packages were used to collect, organize, and save data: *gsread* 5.3.2 (Burnashev *et al.* 2022), *numpy* v1.22.1 (Harris *et al.* 2020), and *pandas* v1.5.2 (McKinney 2010; The pandas Development Team 2022).

Chemical Analysis Methods

The number of features collected about the chemical analyses was too high to input data reliably and directly in a Google or Excel spreadsheet. Instead, I wrote two custom R Shiny apps to facilitate data collection within the working group. One R Shiny app was a form that each member filled with all features listed (Table 4.2, Figure 4.2a). I collected information for GC-FID explicitly because studies often paired quantitative results from runs with different conditions on GC-FID with qualitative results from GC-MS. 529 studies included data from GC-FID. If a study did not analyze concentration data through GC-FID, then the form automatically input ‘NA’ values for those features relevant to GC-FID. While we did not collect the full GC

key	value
<i>chem_method_id</i>	primary key for chemical method
<i>paper_id</i>	secondary key for paper
<i>extraction_type</i>	type of extraction (e.g., dynamic headspace, hydrodistillation)
<i>biomass_init</i>	initial biomass of plant tissue for hydrodistillation or extraction by organic solvent
<i>solvent_compound</i>	name of solvent compound if extracted by organic solvent
<i>fid</i>	binary indicator if study used fid to quantify compound concentration
<i>fid_carrier_gas</i>	carrier gas used for gc-fid
<i>fid_flow_rate</i>	gas flow rate for gc-fid
<i>fid_flow_units</i>	flow rate units for gc-fid
<i>fid_run_time</i>	total run time for gc-fid
<i>fid_inlet_temp</i>	injector temperature for gc-fid
<i>fid_split</i>	use of split method for gc-fid
<i>fid_column_type</i>	capillary column type for gc-fid (e.g., DB-5)
<i>fid_column_length</i>	column length for gc-fid
<i>fid_inner_diameter</i>	inner diameter for gc-fid
<i>fid_film_thickness</i>	film thickness for gc-fid
<i>fid_start_temp</i>	oven start temperature for gc-fid
<i>fid_final_temp</i>	oven final temperature for gc-fid
<i>inlet_temp</i>	injector temperature for gc-ms
<i>carrier_gas</i>	carrier gas used for gc-ms
<i>flow_rate</i>	gas flow rate for gc-ms
<i>flow_units</i>	gas flow units for gc-ms
<i>run_time</i>	total run time for gc-ms
<i>split</i>	use of split method for gc-ms
<i>column_type</i>	capillary column type for gc-ms (e.g., DB-5)
<i>column_length</i>	column length for gc-ms
<i>inner_diameter</i>	inner diameter for gc-ms
<i>film_thickness</i>	film thickness for gc-ms
<i>oven_start_time</i>	oven start temperature for gc-ms
<i>oven_final_temp</i>	oven final temperature for gc-ms
<i>ion_method</i>	ionization method (e.g. electron impact, chemical)
<i>ion_energy</i>	ionization energy (e.g. 70 eV)
<i>ion_source_temp</i>	ionization source temperature

Table 4.2 Structure of the *Chemical methods* table. Blue shading signifies the table's primary key, and yellow signifies the table's secondary key that connects it to other tables.

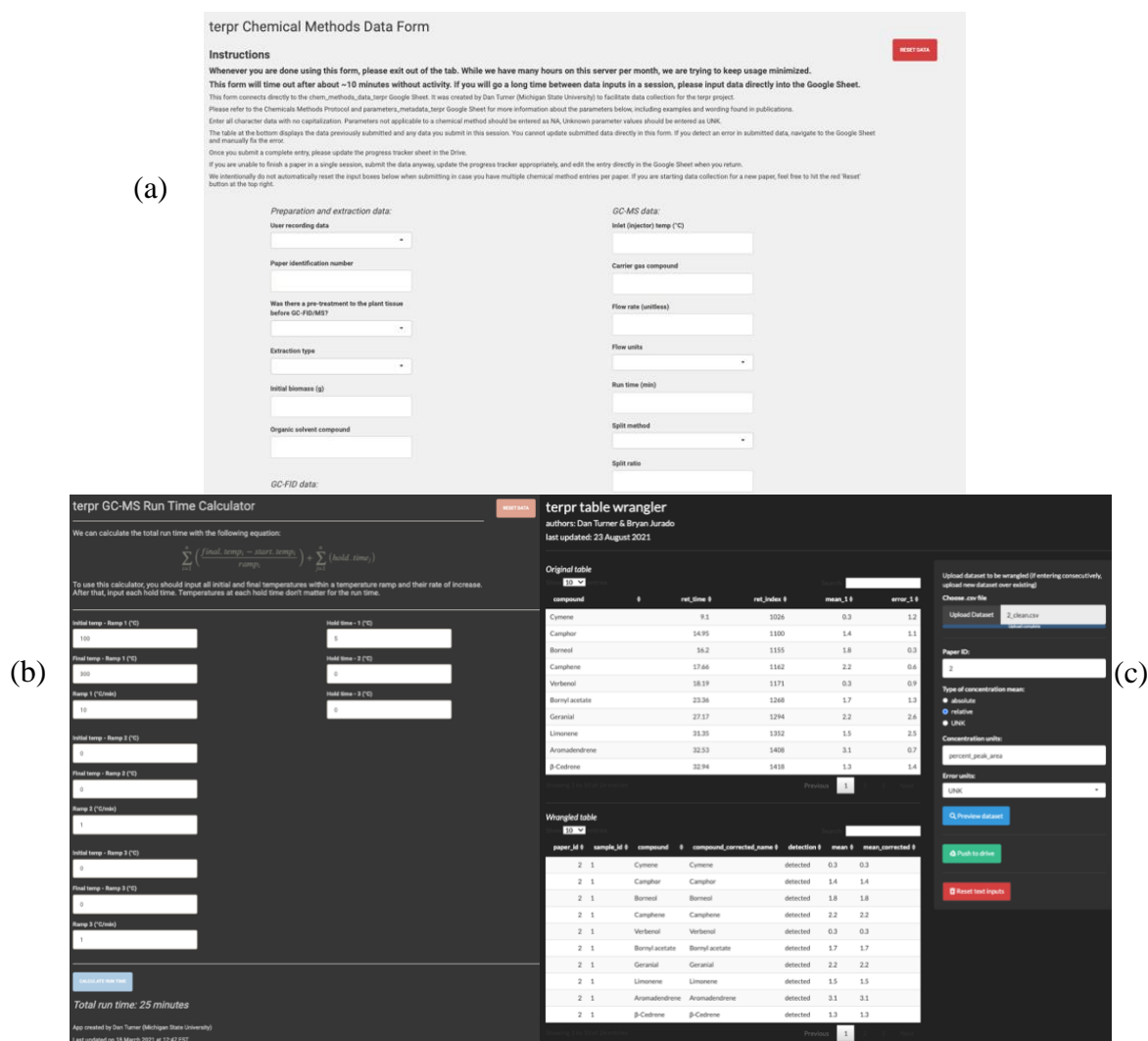


Figure 4.3 Representative previews of Shiny form used (a) to collect chemical methods data, (b) to calculate the machine run time from the temperature program reported in each study, (c) to transform compound and concentration values from each study. Not all fields in the chemical methods data form were included in the final version of *terpr* v1.0.0. After pressing “Calculate Run Time” in (b), the output would be manually copied to the chemical methods form in (a). Transformed (wrangled) tables in (c) were manually inspected for errors and proper formatting. They were then pushed to the cloud for further processing.

temperature program (e.g., ramp, rate of temperature increase, and hold times), we input those values into a “Run Time Calculator” R Shiny app which took those values and output a single run time (Figure 4.3b). The formula for calculating machine run time was:

$$Total\ machine\ run\ time = \sum_{i=1}^n \frac{final.temp_i - start.temp_i}{ramp_i} + \sum_{j=1}^o hold.time_j$$

If the authors reported a final run time in the article that differed from our calculation, then we made a note. If authors used multiple capillary columns with different stationary phases and levels of polarity, we recorded the column as *multiple*, and made a note in the “comments” column. We did not distinguish between push and push/pull dynamic headspace sampling methods. Rather, they are listed as “dynamic headspace.” We did not record if the tissue had been dried before phytochemical extraction. Each sample has only one chemical method, but a study can have multiple chemical methods (e.g., one study had two samples: one extracted with hydrodistillation and one extracted with multiple organic solvents).

Compound names

Some of these steps are identical to those in Chapter 3, but I will reiterate them here. Compound names were copied and pasted from the original article text as a Portable Document Format (PDF) or Hypertext Markup Language (HTML). Compound names were amended by hand if the text was not rendered properly when copying (e.g., original name: α -pinene, copied name: ?-pinene, amended name: a-pinene or alpha-pinene). Greek letters substituted Latin alphabetical spellings in compound names in the custom R Shiny App described in the *Concentration* section below. I manually reviewed every compound name (19543 records without standardizing character cases) to ensure white space, duplicate names, and any overlooked typos from copying and pasting in names were amended.

Studies often presented the same compound with different common names. I standardized these names through either (1) the PubChem application program interface (API) via *PubChemPy v1.0.4* and accepted the first result given (Swain *et al.* 2014; Kim *et al.* 2021) or (2)

for those compounds not resolved with *PubChemPy*, I manually searched PubChem’s website, Wiley SpectraBase, ChemSpider, and the Chemistry WebBook (John Wiley & Sons, Inc. 2022; Linstrom and Mallard 2022; Royal Society of Chemistry 2022). I used regular expressions to convert names between Roman and Greek alphabetical characters in R or Python as needed.

As I standardized the common names, I also collected the Simplified Molecular Input Line System (SMILES) and/or IUPAC International Chemical Identifier Key (InChIKey) to detail the specific configuration and substructures of these compounds if they were available on any of the above sources (Weininger 1988, Heller *et al.* 2015). I automatically collected these values if the compound was identified with *PubChemPy*, and I manually collected these values between May and July 2022 if the compound was found in SpectraBase or ChemSpider. I also wrote a Python script to automatically extract the SMILES and InChIKey from the NIST URL hosting the compound’s information.

Studies reported unidentified compound peaks, unidentified compounds of a known superclass, or compounds of an unidentified configuration. If this occurred, I labeled those compound’s *identification_resolution* as “unknown,” “unknown_{superclass},” or “compound,” respectively. If an “exact” compound match was found (e.g., verbatim name was list in the names of synonyms in PubChem), then the *identification_resolution* was labeled “exact.” In total, I found the Canonical SMILES or InChIKey for 98.0% of all exact compound observations through these four sources. After collecting all these values, each entry was labeled through the *NPClassifier*, via the *NPCTable* function in the *chemodiv* package. *NPClassifier* is an online tool that determines the molecular pathway and class structure of a compound through deep learning algorithms (Kim *et al.* 2020).

key	value
<i>compound_id</i>	primary key for chemical compound
<i>paper_id</i>	secondary key for treatment
<i>sample_id</i>	secondary key for paper
<i>compound_name</i>	Reported name of chemical compound
<i>id_resolution</i>	resolution of compound identification (e.g. exact match, superclass, unknown)
<i>superclass*</i>	classification as monoterpenoid or sesquiterpenoid
<i>smiles**</i>	Canonical SMILES if available
<i>inchikey**</i>	InChIKey if available
<i>detected</i>	chemical compound detection status (e.g., detected, not detected, trace)
<i>concen_mean</i>	mean concentration of chemical compound
<i>concen_error</i>	error around the mean concentration of chemical compound
<i>concen_error_type</i>	type of error value
<i>concen_type</i>	type of concentration value
<i>concen_units</i>	concentration units

Table 4.3 Structure of the *Concentration* table. * signifies that the feature determined with Canonical SMILES and InChIKey with NPClassifier (Kim *et al.* 2020) via the *chemodiv* R package (Petrén *et al.* 2023). ** signifies that the feature determined through the *PubChemPy* Python package or manual searches through PubChem, NIST, SpectraBase, and ChemSpider. Yellow signifies the table's secondary key that connects it to other tables.

Chemical Concentrations

Concentration values were copied and pasted from the original article text simultaneously with the compound names. Each study's table(s) were manually inspected and cleaned of any footnotes and paste errors, and column names were indexed to "pivot" (i.e., transpose) longer in the next step. If a study had more than one table that fit our criteria for inclusion, then we manually concatenated the tables (Figure 4.2). Once the tables in spreadsheets were clean, I deployed a "terpr table wrangler" Shiny app to standardize the form of each table (Figure 4.3c). The Shiny app accepted a .csv file and a required fields about the concentration units and paper

identification number. A button initiated table transformation, a process “pivoting” the table long, adding the additional concentration unit fields, and printing an in-app preview. Given the wide variation across studies in quantitative data reporting styles, the transformation process also standardized the values of each detected, not detected, or trace compound presented in a study. The button converted concentration values of zero, ‘NA’, ‘ND’, ‘-’, blank spaces, and other common indicators to *NA* and listed the value as ‘not_detected’ in the *detection* feature. Values listed as ‘tr’ or ‘trace’ were given *NA* concentration values and listed as ‘trace’ in the *detection* feature. Numerical values greater than zero were listed as ‘detected.’ Some studies only reported presence-absence compound observations, or the tables were already presented in “long” format. Since they fit out of the scope of the Shiny app transformation process, these studies’ tables were transformed by hand in a Google Sheet. All studies’ tables were then concatenated in an R script to complete the *Concentration* table for the database.

The *Concentration* table was joined by compound name to the *Compound* table, and for clarity and ease of use, I present them together in this version of the database (Table 4.3). Only ‘detected’ and ‘trace’ compounds are included. I also excluded all chemicals that were not monoterpenoids or sesquiterpenoids based on the labeling through *NPClassifier*.

Location

Location data were collected mostly from the “Materials and Methods” sections of papers and supplemental materials. Locations were split into two categories: *source* and *experimental* (Table 4.4). Source locations denoted the originating plant tissue that was analyzed for phytochemicals. Experimental locations denoted locations where source plants were relocated, transplanted, or grown in a common garden or agricultural setting. I did not collect location information when plants were grown inside. Source data does not include the

location of seeds used in an experimental, common garden, or agricultural setting. If location data was unclear, then the location coordinates and name was designated “unknown.”

Latitude and longitude coordinates were often presented in Degree Minute Second (DMS) notation in the original study text. I converted all coordinates to decimal degrees at the precision of six decimal places. In 22 locations, the DMS record for either the latitude or longitude of the source or experimental material reported exceeded the maximum possible (e.g., a record claimed the tissue was collected at a latitude of 42°78’41.2N’’). I recorded locality names generally as they were presented in the text with the following format depending on the amount of location information presented: colloquial location name (e.g., Michigan State University), municipality (e.g., East Lansing), region (e.g., Michigan), country (e.g., USA). Each value is delimited by a comma. If a study only presented a country name, I only recorded that country name. However, when a region (e.g., state or province) was reported with no country, I added the country name to the locality information.

key	value
<i>location_id</i>	primary key for location
<i>paper_id</i>	secondary key for paper
<i>exp_long</i>	longitudinal coordinates for common garden location
<i>exp_lat</i>	latitudinal coordinates for common garden location
<i>exp_name</i>	municipality, state/province, country of common garden location
<i>source_long</i>	longitudinal coordinates for wild/source population
<i>source_lat</i>	latitudinal coordinates for wild/source population
<i>source_name</i>	municipality, state/province, country of wild/source population

Table 4.4 Structure of the *Location* table. Blue shading signifies the table’s primary key, and yellow signifies the table’s secondary key that connects it to other tables.

Sample

Data on the identity, sampling effort, phenological stage, and age were collated in the *Sample* table (Table 4.5). Plant genera and species are reported as they were in the papers and are not taxonomically standardized in the *terpr* database. Cultivar and genotype information were recorded as they were written in each study. The *plant_organ* feature is a specific record of the sampled plant part (e.g., anther, pericarp, leaf vein, bark, root) as it was presented in the study. The *organ_mod* feature is a modified from the *plant_organ* feature that has a broad categorization of the type of plant tissue sampled (e.g., floral, fruit, leaf, bark, root). The organ stage was reported verbatim from the study, while the *phenological_stage* feature was inferred or recorded exactly from the text with possible records including “flowering,” “fruiting,” “vegetative,” “sporophyte,” and “unknown.” The *life_stage* feature was inferred or recorded exactly from the text with records such as “seedling,” “sapling,” “juvenile,” “mature.”

I recorded the number of individuals included in a sample, which can be considered biological replicates, in the *num_individuals* feature. I could not independently verify if the number of individuals were genetically unique for all studies. I also recorded the number of technical replicates, runs on the GC-MS or GC-FID, that produced the value (e.g., mean and variance) presented in the concentration table in the *sample_size*. The *data_reported* feature is a brief note about concentration data aggregation in each study, such as a minimum-maximum range (“min_max”), mean and variance (“mean_error”), mean, and solely presence/absence data (“presence_absence”).

key	value
<i>sample_id</i>	primary key for sample
<i>paper_id</i>	secondary key for paper
<i>treatment_id</i>	secondary key for treatment
<i>chem_method_id</i>	secondary key for chemical method
<i>location_id</i>	secondary key for location
<i>family</i>	plant family name
<i>genus</i>	plant genus name
<i>species</i>	plant species name
<i>subspecies</i>	plant subspecies name
<i>genotype</i>	plant genotype/variety
<i>life_stage</i>	plant life stage
<i>organ_stage</i>	organ stage
<i>plant_age</i>	plant age
<i>tissue_age</i>	tissue age
<i>num_individuals</i>	number of plant individuals (biological replicates) represented in sample
<i>plant_organ</i>	plant organ sampled
<i>sample_size</i>	number of technical replicates for reported concentration
<i>data_reported</i>	aggregation factor for data (e.g., mean, mean_error, median, presence/absence)

Table 4.5 Structure of the *Sample* table. Blue shading signifies the table's primary key, and yellow signifies the table's secondary key that connects it to other tables.

Study

The study table includes all results from the searches, inclusive of those papers that did not fit the criteria for collecting all data (Table 4.6). I include it here in case other researchers would like to examine those studies that were excluded or to collect data from those studies for which I did not have the resources to collect all data. The following fields were based on the output from the export of the search term results on ISI Web of Science: *authors*, *title*, *journal*, *pub_date*, *pub_year*, *volume*, *issue*, *start_page*, *end_page*, and *doi*.

key	value
<i>paper_id</i>	primary key for paper
<i>authors</i>	list of study's authors
<i>title</i>	study title
<i>journal</i>	study publication
<i>pub_date</i>	publication date of study
<i>pub_year</i>	publication year of study
<i>volume</i>	study volume
<i>issue</i>	study issue
<i>start_page</i>	first page of issue that contains study
<i>end_page</i>	last page of issue that contains study
<i>search_terms</i>	the search terms on ISI web of science that resulted in this study
<i>doi</i>	digital object identifier
<i>exclusion</i>	indicator for exclusion from the database (values: 0 – included in database; 1 – did not fit title criteria; 2 – did not fit main text criteria; 3 – removed due to limited resources to process; 4 – could not locate text or required explicit requests to authors; 5 – duplicated across searches)

Table 4.6 Structure of the *Study* table. Blue shading signifies the table's primary key.

Treatment

Data on general plant growth conditions and special experimental treatments were collected for each study (Table 4.7). The *collection_year* feature is noted for all plants grown outside, regardless of cultivated or wild growth. If collection occurred across multiple years, and the samples years were not disaggregated in the concentration table, then all years were recorded and delimited by commas. I collected the *location_type* feature as eight possible factors: (1) “common_garden” for plants grown on farms, experimental plots planted by humans, botanical gardens, arboreta, (2) “wild_population” for plants grown in the wild, agnostic to native and cultivated statuses, (3) “growth_chamber” for plants grown in small chambers under controlled conditions, (4) “commercial” for plant material bought as an essential oil or in a store or market with unknown growth conditions, (5) “greenhouse” for plants grown in rooms under controlled conditions, (6) “screenhouse” for plants grown in rooms with exposure to natural weather

conditions, but excluded from wild ecological interactions, (7) “unknown” for unknown growth location type, and (8) “multiple” for plant tissue sampled across multiple location types.

Treatment features, such water, thermal, and nutrient (i.e., fertilizer) stress, herbivory damage, and pathogen inoculation, were indicated by “yes” or “no.” A “yes” under these treatment features indicates the sample had the treatment. A “no” under these treatment features indicates the sample did not have the treatment. Some studies resampled plants and/or plant communities multiple times after a treatment, and I collected information on the time since the treatment was applied in days. For example, a value of “-1” indicates that the sample was collected one day before a treatment was applied, and a value of 365 indicates that the sample was collected a year after a treatment was applied. If the *time_since_treatment* feature has a numeric value, but there is not a “yes” indicated in a treatment feature, then the study may have a treatment that is not within the scope of the data we collected.

key	value
<i>treatment_id</i>	primary key for treatment
<i>paper_id</i>	secondary key for paper
<i>fertilizer_treatment</i>	indicator if fertilizer treatment
<i>water_treatment</i>	indicator if water treatment
<i>temp_reported</i>	indicator if growing temperature was recorded
<i>pressure_reported</i>	indicator if ambient pressure was recorded
<i>temp_treatment</i>	indicator if temperature treatment
<i>fire_treatment</i>	indicator if fire treatment
<i>time_measure_treatment</i>	time since treatment
<i>herbivory_treatment</i>	indicator if herbivory was a treatment
<i>pathogen_treatment</i>	indicator if pathogen inoculation/introduction was a treatment
<i>location_type</i>	type of growth conditions (e.g., wild population, greenhouse)
<i>collection_year</i>	year of tissue collection if grown outside

Table 4.7 Structure of the *Treatment* table. Blue shading signifies the table’s primary key, and yellow signifies the table’s secondary key that connects it to other tables.

Data Records

Access

The *terpr v1.0.0* database can be accessed as .csv files upon request to the author.

Development and deployment as an SQL database are ongoing.

Data coverage

Overall, the database includes 1227 studies containing 5107 samples from 1178 plant species. The species counts are based on standardizing names with the Leipzig catalog of vascular plants (Freiberg *et al.* 2020), and non-vascular plant names were not standardized for the count. The database has high geographical coverage in Europe, Northern Africa, and South

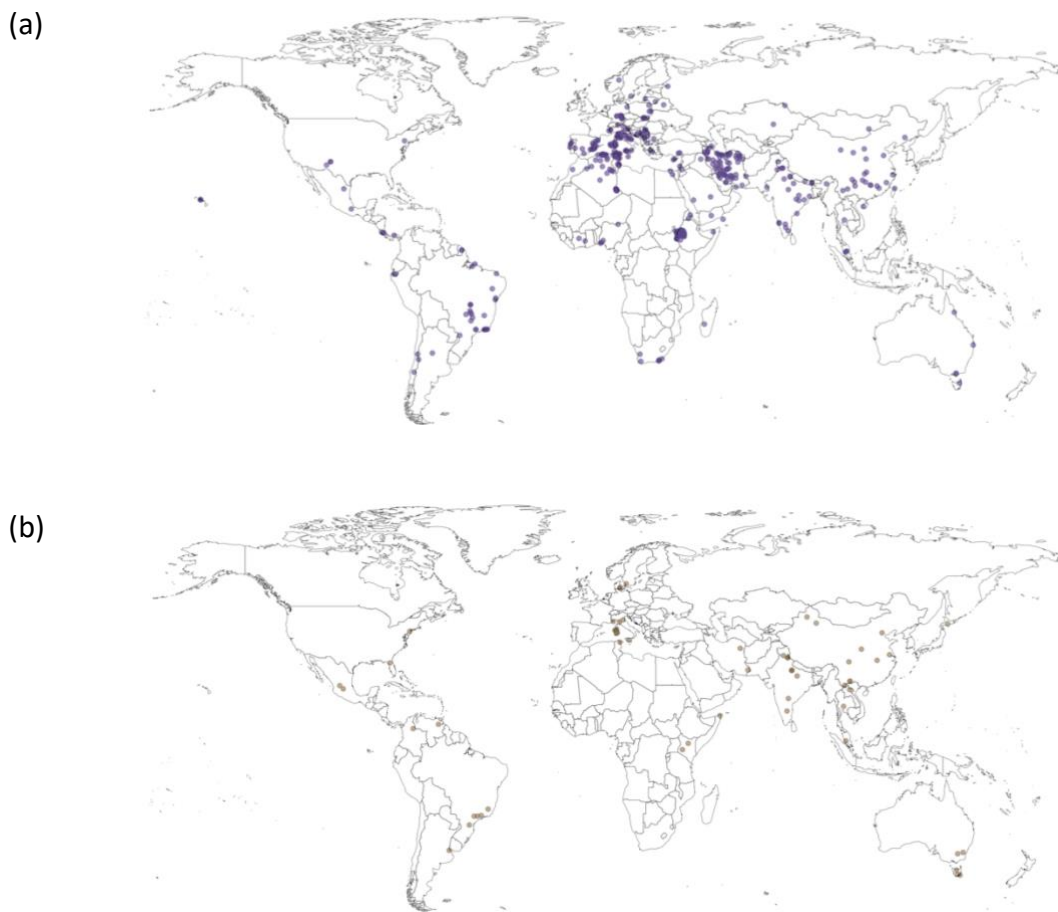


Figure 4.4 Locations for (a) wild-grown plants and (b) plants grown in outdoor common garden settings.

Asia (Figure 4.4). Few samples are from North America, the Amazon, North Asia, Central Africa, and Western Australia. The largest coverage within angiosperm clades is in the Lamiales

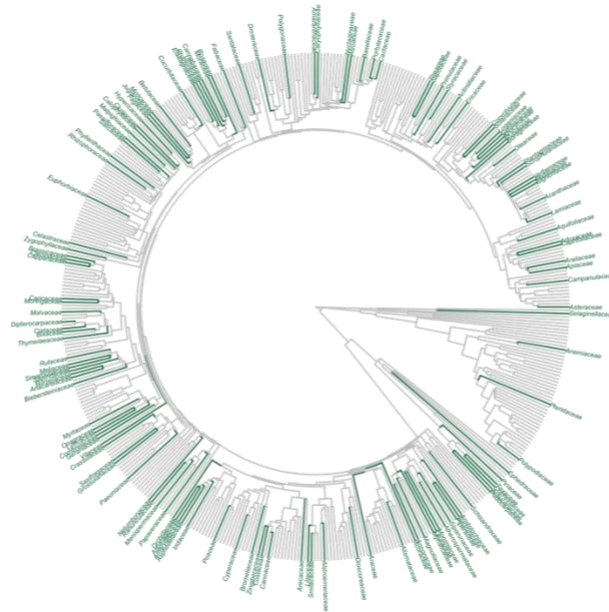


Figure 4.5 Vascular plant phylogeny of all plant families. Green and labeled tips indicate families present in the database. The plant phylogeny is based on the GTOTB.extended tree of vascular plant families (Smith and Brown 2018). All family names were standardized using the *lcvplants* R package (Freiberg *et al.* 2020).

(number of species = 197) and Asterales (number of species = 154), and high representation in Pinales (number of species = 77; Figure 4.5). Ferns and non-vascular plants do not have high sampling in the database, representing less than 5% of all plants species in the database,

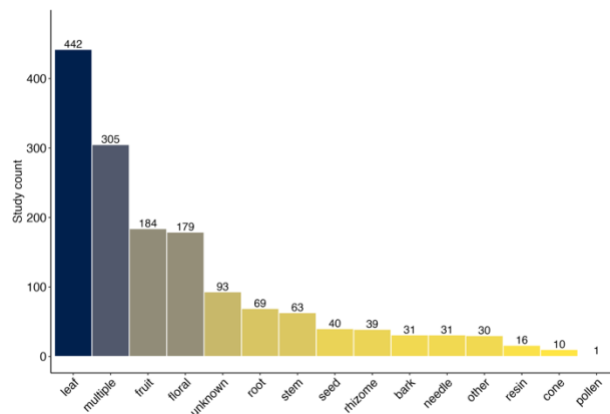


Figure 4.6 Study count by grouped organ feature (*Sample* table feature: “*organ_mod*”).

respectively. The database is biased toward foliar tissue samples, followed by the grouped aerial

parts of plant tissue (i.e., shoots), which contained stem, foliar tissue, and sometimes fruit or floral tissue depending on the sampled phenological stage (Figure 4.6). The mean machine run time across all chemical methods was 56.3 minutes \pm 23.5 (1SD). The most common extraction type was hydrodistillation (Figure 4.7).

Accounting for configuration, charge, and other structural differences between

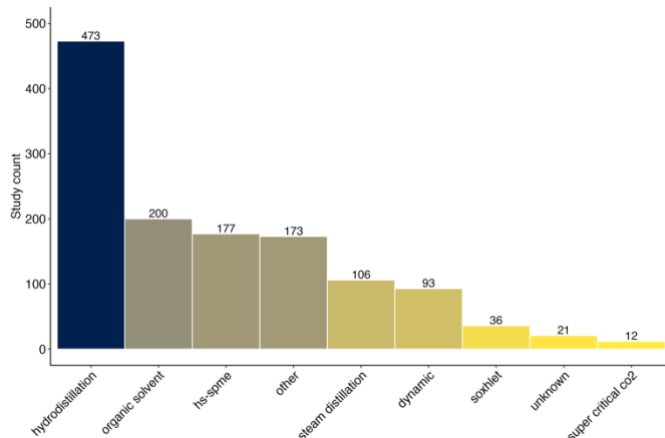


Figure 4.6 Study count by extraction type feature (*Chemical methods* table feature: “*extraction_type*”).

compounds, 1852 unique monoterpenoids and sesquiterpenoids are reported across all studies’

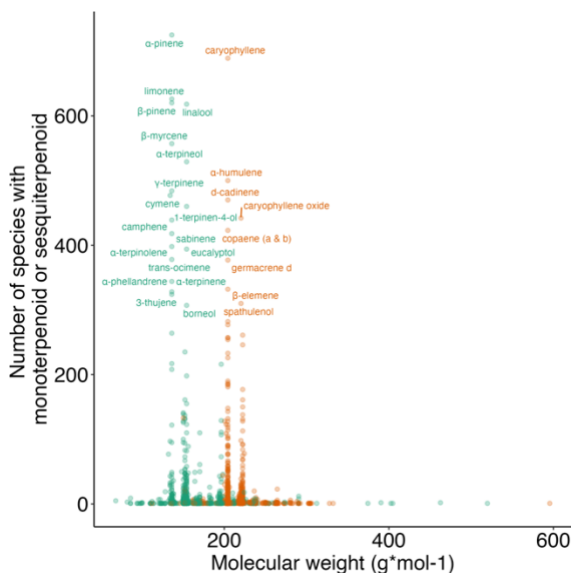


Figure 4.7 Number of species with observed monoterpenoid or sesquiterpenoid. Superclass was determined by *NPClassifier* (Kim *et al.* 2020) and molecular weight was collected from PubChem, NIST, ChemSpider, or SpectraBase.

6829 total unique compounds of all classes. A few compounds were more present across the

species in my database than others. In terms of species coverage, α -pinene and caryophyllene were the most commonly found monoterpene and sesquiterpene, respectively (Figure 4.7).

Technical Validation

Several data quality, collection process, and transformation checks were incorporated into the workflow. First, whenever possible, data collection was automated to reduce mistakes (i.e., transforming wide table formats to long table formats through the R Shiny app). Despite reduced time and energy efficiency, paste errors from concentration data collection were corrected manually to ensure data integrity. I validated all compound names for chemical metadata across at least one of the four sources—PubChem, ChemSpider, SpectraBase, and NIST.

Beyond its necessity for the database structure, sample indexing was a detailed validation process ensuring that the secondary keys in the *Sample* table were properly related to the primary keys of the *Treatment*, *Location*, and *Chemical methods* tables. Locations were also validated via visual inspection on a map. Finally, a random sample of 50 samples were chosen for validation with the *Concentration* table, with an error rate of 2% ($n = 1$ sample). In this case, the sample had the incorrect cultivar listed, and it was corrected. With such a low error rate, it would not be time or cost efficient to validate more samples. However, users are welcome to validate, check, and provide feedback to enhance the integrity of such a large database.

Usage Notes

While I attempted to achieve both breadth and depth in this database, several limitations persist when compiling data across many sources, including information that was not reported or was outdated. The *terpr v1.0.0* was designed with this fact in mind, and I have built in several features to connect to other databases and software packages to overcome reporting limitations, changes in taxonomy, and chemical classification (Figure 4.2). However, due to resource

limitations, I was unable to record the retention indices and times and their associated library citations for identification. A future iteration of the database may include these data from each study, if available.

I provide the Canonical SMILES and InChIKey values for connection to various chemical databases, including *NPClassifier*, PubChem, and ChemSpider. Chemical metadata and properties can be connected to the database through various chemical database APIs. To facilitate these connections, users can deploy existing software development kits that exist for Python, such as *PubChemPy* and *ChemSciPy* (Swain 2018), in R, such *chemodiv*, *webchem* (Szöcs *et al.* 2020), and *ChemmineR* (Cao *et al.* 2008).

For the location data, I recommend the *tidygeocoder* R package (Cambon *et al.* 2021) to supplement the latitude and longitude coordinates supplied here. The *tidygeocoder* R package contains functions that can search municipality names across various geolocation APIs and append latitude and longitude data. I recommend validating and standardizing plant taxonomy via the *lcvplants* R package, or equivalent taxonomic standardization software package, while following the protocol outlined in Grenié *et al.* (2021), as I did to complete Figure 4.4 (Freiberg *et al.* 2020).

While not a completely exhaustive collection of studies, the database can be a starting point or supplement for meta-analyses into the effects of these treatments on terpenoid emissions and diversity. However, the coverage across treatment features varies widely. For example, 47 studies have herbivory treatments, while eight studies have pathogen inoculation treatments. Other search criteria for treatments may yield more studies, similar to the targeted searches I did for studies with herbivory treatments. Additionally, beyond being a reference for previously successful chemical identification methods, database meta-analyses can determine the efficacy of

various capillary columns, machine run times, extraction times, and other detection parameters collated here.

Finally, this database can be a resource to answer foundational questions across, but not exclusive to, four disciplines: *ecology* (e.g., are specific terpenoids associated with particular species interactions? Do plants that produce more diverse terpenoid profiles interact with more species? Can we identify multifunctionalities in particular compounds? Are there synergistic effects on plant community interactions when producing more structurally diverse terpenoid profiles?), *evolution* (e.g., are there hotspots of terpenoid diversity across the plant phylogeny? How can we leverage phylogenetic comparative techniques to study terpenoid evolution? Is the origin of the production of common monoterpenoids and sesquiterpenoids common across all groups or did it evolve independently?), *conservation* (e.g., are threatened areas more prone to chemical diversity loss in addition to biological diversity loss?), and *application* (e.g., are there synergistic effects when combining terpenoids for medicinal use? Are there taxonomic reservoirs of untapped terpenoid diversity for applied exploration?).

Code Availability

All code to demonstrate the core processes of the workflow (Figure 4.2) is available upon request to the author. Additionally, the code to reproduce the figures here is available upon request. Intermediary tables (e.g., individual studies' wide-formatted and cleaned concentration tables) are also available upon request.

REFERENCES

- Becerra JX (1997) Insects on plants: Macroevolutionary chemical trends in host use. *Science* 276:253–256. <https://doi.org/10.1126/science.276.5310.253>
- Burnashev A (2022) gspread. <https://github.com/burnash/gspread/>
- Cambon J, Hernangómez D, Belanger C, Possenriede D (2021) tidygeocoder: An R package for geocoding. *Journal of Open Source Software* 6:3544. <https://doi.org/10.21105/joss.03544>
- Cao Y, Charisi A, Cheng L-C, et al (2008) ChemmineR: a compound mining framework for R. *Bioinformatics* 24:1733–1734. <https://doi.org/10.1093/bioinformatics/btn307>
- Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* 66:212–229. <https://doi.org/10.1111/j.1365-313X.2011.04520.x>
- Duke J, Bogenschutz MJ (1994) Dr. Duke's phytochemical and ethnobotanical databases. USDA, Agricultural Research Service, Washington, DC
- Freiberg M, Winter M, Gentile A, et al (2020) LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Scientific Data* 7:416. <https://doi.org/10.1038/s41597-020-00702-z>
- Garnier S, Ross N, Rudis R, et al (2021) Rvision - Colorblind-Friendly Color Maps for R. <https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html>
- Grenié M, Berti E, Carvajal-Quintero J, et al (2023) Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. *Methods in Ecology and Evolution* 14:12-25 <https://doi.org/10.1111/2041-210X.13802>
- Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry* 68:2831–2846. <https://doi.org/10.1016/j.phytochem.2007.09.017>
- Heller SR, McNaught A, Pletnev I, et al (2015) InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics* 7:23. <https://doi.org/10.1186/s13321-015-0068-4>
- Howe AD, Costanzo M, Fey P, et al (2008) Big data: The future of biocuration. *Nature* 455:47–50. <https://doi.org/10.1038/455047a>
- International Society for Biocuration (2018) Biocuration: Distilling data into knowledge. *PLOS Biology* 16:e2002846. <https://doi.org/10.1371/journal.pbio.2002846>
- Jacobsen A, de Miranda Azevedo R, Juty N, et al (2020) FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence* 2:10–29. https://doi.org/10.1162/dint_r_00024

- John Wiley & Sons, Inc. (2022) SpectraBase. <https://spectrabase.com/>
- Junker RR, Kuppler J, Amo L, et al (2018) Covariation and phenotypic integration in chemical communication displays: biosynthetic constraints and eco-evolutionary implications. *New Phytol* 220:739–749. <https://doi.org/10.1111/nph.14505>
- Kalske A, Shiojiri K, Uesugi A, et al (2019) Insect Herbivory Selects for Volatile-Mediated Plant-Plant Communication. *Current Biology* 29:3128–3133.e3. <https://doi.org/10.1016/j.cub.2019.08.011>
- Kessler A, Kalske A (2018) Plant secondary metabolite diversity and species interactions. *Annu Rev Ecol Evol Syst* 49:115–138. <https://doi.org/10.1146/annurev-ecolsys-110617-062406>
- Kim H, Wang M, Leber C, et al (2020) NPClassifier: A deep neural network-based structural classification tool for natural products. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.12885494.v1>
- Kim S, Chen J, Cheng T, et al (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* 49:D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>
- Kratochvil M (2022) scattermore: Scatterplots with more points. <https://github.com/esaexa/scattermore>
- Linstrom P, Mallard W (2022) NIST Chemistry WebBook, NIST Standard Reference Database Number 69. National Institute of Standards and Technology, Gaithersburg, Maryland
- Loub WD, Farnsworth NR, Soejarto DD, Quinn ML (1985) NAPRALERT: Computer handling of natural product research data. *J Chem Inf Comput Sci* 25:99–103. <https://doi.org/10.1021/ci00046a009>
- McKinney W (2010) Data Structures for Statistical Computing in Python. Austin, Texas, pp 56–61
- Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, et al (2018) IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Sci Rep* 8:4329. <https://doi.org/10.1038/s41598-018-22631-z>
- Müller K, Wickham H (2021) tibble: Simple Data Frames. <https://tibble.tidyverse.org/>
- R Core Team (2021) R: A language and environment for statistical computing. <https://www.r-project.org/>
- Royal Society of Chemistry (2022) ChemSpider. <http://www.chemspider.com/>. Accessed 2022

- Sansone S-A, McQuilton P, Rocca-Serra P, et al (2019) FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol* 37:350–369. <https://doi.org/10.1038/s41587-019-0080-8>
- Slowikowski K (2021) ggrepel: Automatically position non-overlapping text labels with “ggplot2”. <https://CRAN.R-project.org/package=ggrepel>
- Smith SA, Brown JW (2018) Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany* 105:302–314. <https://doi.org/10.1002/ajb2.1019>
- Sorokina M, Steinbeck C (2020) Review on natural products databases: where to find data in 2020. *J Cheminform* 12:20. <https://doi.org/10.1186/s13321-020-00424-9>
- Swain M (2018) ChemSpiPy: A Python wrapper for the ChemSpider API. <https://github.com/mcs07/ChemSpiPy>
- Swain M, Kurniawan E, Powers Z, et al (2014) PubChemPy. <https://pubchempy.readthedocs.io/en/latest/>
- Szöcs E, Stirling T, Scott ER, et al (2020) webchem: An R package to retrieve chemical information from the web. *J Stat Soft* 93(13). <https://doi.org/10.18637/jss.v093.i13>
- Tetali SD (2019) Terpenes and isoprenoids: a wealth of compounds for global use. *Planta* 249:1– 8. <https://doi.org/10.1007/s00425-018-3056-x>
- The pandas development team (2022) pandas-dev/pandas: Pandas
- Tholl D (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Current Opinion in Plant Biology* 9:297–304. <https://doi.org/10.1016/j.pbi.2006.03.014>
- Ushey K, Allaire J, Tang Y (2022) reticulate: Interface to “Python”. <https://rstudio.github.io/reticulate/>
- Weber J, Archer-Nicholls S, Abraham NL, et al (2022) Chemistry-driven changes strongly influence climate forcing from vegetation emissions. *Nat Commun* 13:7202. <https://doi.org/10.1038/s41467-022-34944-9>
- Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Model* 28:31–36. <https://doi.org/10.1021/ci00057a005>
- Wickham H (2016) *Elegant Graphics for Data Analysis*. Springer-Verlag New York
- Wickham H (2020) httr: Tools for working with URLs and HTTP. <https://httr.r-lib.org/>
- Wickham H (2019) stringr: Simple, consistent wrappers for common string operations. <https://stringr.tidyverse.org/>

Wickham H, François R, Henry L, Müller K (2022) dplyr: A grammar of data manipulation.
<https://dplyr.tidyverse.org/>

Wickham H, Girlich M (2022) tidyr: Tidy messy data. <https://tidyr.tidyverse.org/>

Zu P, Boege K, del-Val E, et al (2020) Information arms race explains plant-herbivore chemical communication in ecological communities. *Science* 368:1377–1381.
<https://doi.org/10.1126/science.aba2965>

CHAPTER 5:

CONCLUSION

As is often the case in a doctoral program, this dissertation journey had many twists and turns before settling on its final state here. However, half of the duration of my dissertation occurred during the COVID-19 pandemic and its related shutdowns. Most of my experimental plants died in the greenhouse in March and April 2020 because I could not access them, I lost all hired labor resources in the 2020 field season, and I was unsure when or if processes, resources, places, and supply chains would return to their previously functional levels. I decided to shift the substance of my dissertation from primarily field-based ecological experiments on plant-insect interactions to computer-based, large data work on phytochemical diversity. By addressing the pandemic's effects on this dissertation, I rationalize the seemingly disparate subject matters between the goldenrod chapter and the terpenoid chapters. For each project, I address room for growth, methods that worked well, and some that did not work so well.

Temporal context of herbivory affects goldenrod community ecology and plant growth

Chapter 1 focused on the community ecology of a plant and its antagonists. I found that the temporal context following a species interaction event (e.g., early-season mirid feeding or variation in jasmonic acid spray timing) had significant influences on the conclusions deduced from the results. These results emphasize the necessity for temporally explicit ecological experiments, a perspective that is growing in ecology (Yang 2020). We can look to the successes and failures of my experimental approaches to arrive at a better understanding of temporal dynamics in the natural world.

My common garden and *in situ* field experiments evaluated intra-annual variation in species interactions and plant growth at various time points throughout a single growing season. Ecological systems can exhibit immediate or lagged responses to events that vary in duration and

intensity (Jackson *et al.* 2021). These lagged responses have not been thoroughly investigated in an evolutionary perspective. Therefore, a core question in the evolutionary ecology of plant-insect interactions will be answered if “immediate” or “lagged” responses result in greater fitness. For instance, in my common garden experiment, mirid-fed plants exhibited reduced pathogen loads earlier in the season, but not later in the season. Future experiments may address the fitness benefits of that more “immediate” response to an herbivory event, rather than a lagged or sustained response. Another next step would be replicating previous ecological experiments and measuring responses at different time points along an organism’s ontogeny or phenology to determine if the observed effects vary temporally.

The biology and ecology of *S. altissima* and its herbivores are conducive for experimental manipulation, given my experience with the approaches I took. The application of herbivory feeding by the specialist mirid, *Slaterocoris* sp., was generally straightforward. The white stippling created as the bug feeds on the leaf mesophyll is clearly visible by eye and thus allows for the standardization of herbivory application. Beyond the mirid-feeding application, I also recommend measuring rhizomes by volume displacement as opposed to length or weight if clonally propagating *S. altissima*. From my personal observation, different putative *S. altissima* genotypes varied widely in rhizome thickness; some putative genotypes were at least half as thick than other putative genotypes. Rhizome cut closer to the aboveground shoot also tended to be thicker than rhizomatous tissue farther from the aboveground shoot. Rhizomes can also dry out if not kept moist in a refrigerated environment before replanting. Since they are wet, the mass cannot be measured reliably. Although a more tedious and time-consuming method, volume displacement avoids these biases in length- and mass-based measurements.

To address some of the issues that I had in these experimental protocols, I would avoid pulling rhizomes in the spring. In southwestern Michigan, by the time I had pulled rhizomes in April and May, the wild-growing plants may have already received several bouts of warm weather to begin budding fresh shoots. By placing the plants in the greenhouse after the spring cues had already arrived, dozens of potential ramets never sprouted from the soil. The following year, when I had planted rhizomes that were dug up in December and sat in a refrigerator near 5°C for a month, most rhizomes sprouted. I had little success trying to induce sprouting via bagging rhizomes with wet paper towels and placing in the greenhouse. I do not recommend this protocol without carefully monitoring the humidity inside the bag and avoiding burning the rhizomes.

Finally, I initially planned to run field and greenhouse experiments based on intraspecific diversity in goldenrod's responses to species interaction events. However, this was not possible because my collaborators were unable to genotype the plants by microsatellite amplification after troubleshooting primer and DNA sets, combinations, and thermal cyclers. I do not recommend combined molecular DNA and experimental approaches without some advances in our ability to genotype this species and its complex.

Plant terpenoid diversity varies with temperature, precipitation, and phylogeny: a meta-analysis

I aimed for the first analysis of the terpenoid database to address large, synthetic questions about how the diversity of these compounds varies across plants. In a practical sense, I chose climatic and phylogenetic covariates because questions about their relationship with phytochemical diversity required minimal additional data collection beyond the 100,000s of data points I had already collected for the database. The two climate variables I included—annual mean temperature and annual total precipitation—are only two covariates that could influence

monoterpenoid and sesquiterpenoid diversity. As mentioned in the meta-analysis chapter, a crucial next step in this research is observing trends across gradients of atmospheric pressure. I could not easily locate and extract this covariate with publicly available datasets, but that does not mean it cannot be obtained in the future or that other approaches will fail. Within the WorldClim data, information on temperature variability, seasonality, and solar radiation could be incorporated into complex models for a more complete understanding of the abiotic correlates of plant monoterpenoids and sesquiterpenoids. These models would need to be carefully crafted, however, to consider the expected covariation between climatic variables.

The tools and workflows I employed for data collection here only scratched the surface of available analyses. To calculate β -diversity, I had to pull the individual PubChem fingerprints for all 881 compound features that make each unique. My analyses were agnostic to the specific identity of these features, but we should not always limit our analyses like this. For example, if those compounds that have more hydroxyl groups are more hydrophilic and require greater stomatal conductance to be released from plant tissue, then we could explore the prevalence of these molecular features across climate gradients. We can also incorporate compound properties that are relevant for ecological interactions, such as molecular weight, into our analyses. With our advancements in tools and frameworks for collecting compound metadata and properties, investigations on the perception and effects of phytochemicals on plant-insect interactions can synthesize my approaches developed in this dissertation with experiments to identify mechanisms in field, greenhouse, and lab settings.

When analyzing this amount of data especially in Bayesian contexts, models can take a long time to converge. I found that simply centering and scaling my continuous variables for temperature and precipitation, so they were within the same order of magnitude and centered

around their respective mean, vastly decreased model duration and improved model convergence. Additionally, I was not able to confidently determine phylogenetic signal in α -diversity by calculating Pascal's λ from the models' group-level effects. I recommend future analyses run more complex comparative analyses to find the appropriate measure for phylogenetic signal (Münkemüller *et al.* 2012).

No meta-analysis encompasses every bit of information available to answer a set of questions, and this meta-analysis is no exception. Several limitations arose regarding location data availability. Many studies were excluded from the analyses in this chapter because the authors did not provide precise enough geographical information or were unclear if the plants were growing in a wild or cultivated context. Fortunately, I could increase the sample size of those studies with latitude and longitude coordinates by deploying the *tidygeocoder* R package to those samples with municipality-level location descriptions. I also found that the WorldClim data had wide coverage across the globe for these types of analyses; only twelve samples that had geographic coordinates were not matched to climate data in WorldClim. While data quality will always be vital to the success of any meta-analysis (Koricheva and Gurevitch 2014), context-appropriate database connections can facilitate increasing sample sizes when inclusion criteria are strict and lengthy.

terpr v1.0.0: A database of plant monoterpenoids and sesquiterpenoids

Assembling the terpenoid database required more manual labor than anticipated. Each study required meticulous care to collect its dozens of variables along with chemical compound tables containing, sometimes, hundreds of data points. Programmatic extraction of compound concentration tables from PDFs was not possible for several reasons. For tables that were embedded as text in rows and columns, papers reported the tables too variably (i.e., the table

names and captions were worded too differently or placed in different sequences by study) to automate extraction. Despite growing calls for standard table formats (Broman and Woo 2018), many tables were not rectangular, contained implicit column headers, and empty cells, which required a human to review and fix inconsistencies. For tables in PDFs that are scans of documents, tools, such as optimal character recognition (OCR), would be required for table extraction, which is computationally intensive, and I did not have the resources or time to run tests on the accuracy of OCR methods for the purposes of creating this database. Since most papers were published since the proliferation of digitally published academic articles, these cases were not as common as the others.

If I were to replicate this project, I would hire a team of about a dozen researchers with expertise in analytical chemistry, plant biology, terpenoids, and meta-analyses. I would have two separate individuals collect data on each paper, compare their values, and discuss to arrive at a mutually agreeable conclusion for each database feature. Early in the data collection process, I had to remove a table that recorded the retention index, time, and specification about compound identification (e.g., reference mass-spectrometry library) due to time and resource constraints. If future researchers would like to prioritize collecting more data from the papers already existing in the database, I highly recommend spending resources collecting this information.

A few technical tools were vital in my work with the terpenoid database. Importantly, not all tools are available in one programming language, and not all tools had equal efficiency. My preferences for some software development kits and packages for certain tasks, such as the Python library PubChemPy (Swain *et al.* 2014) for compound metadata searches and the R package *lcvp* (Freiberg *et al.* 2020) for reconciling plant species names, do not mean these are the only tools available and it is possible that others will be developed in the future. Switching

between languages was challenging logistically, but “newer” tools, such as the *reticulate* R package (Ushey *et al.* 2022) and Quarto documents created by Posit Software, present promising avenues to integrate code between programming languages within the same pipelines.

Unfortunately, the data collection process for a database like *terpr* cannot be streamlined into a single script (e.g., .R or .py file) or notebook (e.g., Quarto, RMarkdown, or Jupyter Notebook). A likely solution would be to integrate those data orchestration tools used in applied and business contexts (e.g., Prefect or Airflow). Most emphasis on inter-database connections with my monoterpenoid and sesquiterpenoid database was on the relationships with other chemical databases (e.g., those with ChemSpider and PubChem; Kim *et al.* 2021, Royal Society of Chemistry 2022), but connection to plant trait databases, like TRY (Kattge *et al.* 2020), FunAndes (Báez *et al.* 2022), and the China Plant Trait Database (Wang *et al.* 2022), can be similarly valuable.

If future researchers would like to collect data from newly published papers, I would invite authors to a portal with prompts and file submission guidelines for including data in the database. Submitted data can then be cross-checked by a trained individual on the database team. I wrote custom Shiny apps to facilitate the data collection, which can offer a start to the data submission portal. However, before that, a reasonable next step would be to return to those papers that I excluded due to resource constraints. I also only included monoterpenoids and sesquiterpenoids in this version of *terpr*, but many chemical “byproducts” of the data pipeline (e.g., green leaf volatiles and fatty acids) were observed and excluded and could thus be incorporated into other databases or analyses. If desired, the database could be augmented to all studies that analyze plant tissue with gas-chromatography mass-spectrometry, regardless of focus phytochemical superclass. Finally, chemical ecologists and biochemists alike can apply the many

tools and scripts that I provide in this chapter to chemicals found in the tissue of other organisms, such as insects, fungi, and algae.

REFERENCES

- Báez S, Cayuela L, Macía MJ, *et al.* 2022. FunAndes – A functional trait database of Andean plants. *Sci Data* **9**: 511.
- Broman KW and Woo KH. 2018. Data organization in spreadsheets. *The American Statistician* **72**: 2–10.
- Freiberg M, Winter M, Gentile A, *et al.* 2020. LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Scientific Data* **7**: 416.
- Jackson MC, Pawar S, and Woodward G. 2021. The temporal dynamics of multiple stressor effects: From individuals to cosystems. *Trends in Ecology & Evolution* **36**: 402–10.
- Kattge J, Bönisch G, Díaz S, *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* **26**: 119–88.
- Kim S, Chen J, Cheng T, *et al.* 2021. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **49**: D1388–95.
- Koricheva J and Gurevitch J. 2014. Uses and misuses of meta-analysis in plant ecology. *Journal of Ecology* **102**: 828–44.
- Münkemüller T, Lavergne S, Bzeznik B, *et al.* 2012. How to measure and test phylogenetic signal. *Methods in Ecology and Evolution* **3**: 743–56.
- Royal Society of Chemistry. 2022. ChemSpider.
- Swain M, Kurniawan E, Powers Z, *et al.* 2014. PubChemPy.
- Ushey K, Allaire J, and Tang Y. 2022. reticulate: Interface to “Python.”
- Wang H, Harrison SP, Li M, *et al.* 2022. The China plant trait database version 2. *Sci Data* **9**: 769.
- Yang LH. 2020. Toward a more temporally explicit framework for community ecology. *Ecological Research* **35**: 445–62.

APPENDIX 1:
RECORD OF DEPOSITION OF VOUCHER SPECIMENS

The specimens listed below have been deposited in the named museum as samples of those species or other taxa, which were used in this research. Voucher recognition labels bearing the voucher number have been attached or included in fluid preserved specimens.

Voucher Number: 2023-03

Author and Title of thesis:

Daniel B. Turner, "Time and terpenoids: experimental and data-intensive investigations into temporal ecology and phytochemistry"

Museum(s) where deposited:

Albert J. Cook Arthropod Research Collection, Michigan State University (MSU)

Specimens:

Family	Genus-species	Life Stage	Quantity	Preservation
Miridae	<i>Slaterocoris</i> sp.	Adult	10	point pinned

APPENDIX 2:

SUPPLEMENT FOR TEMPORAL CONTEXT OF HERBIVORY AFFECTS

GOLDENROD COMMUNITY ECOLOGY AND PLANT GROWTH

Putative genotype	Source latitude	Source longitude	Date collected	Dates rhizome planted in greenhouse
152	42.4903847	-85.448568	10-May-19	10-May-19
153	42.4769322	-85.461359	10-May-19	10-May-19
163	42.4845879	-85.451522	12-Apr-19	25-Apr-19
171	42.4841886	-85.451529	19-Apr-19	25-Apr-19
180	42.487991	-85.448502	09-May-19	10-May-19
185	42.4849231	-85.451122	19-Apr-19	25-Apr-19
197	42.4771473	-85.460672	10-May-19	10-May-19
220	42.4904796	-85.448968	10-May-19	10-May-19
242	42.4878199	-85.448148	19-Apr-19	25-Apr-19
308	42.4899983	-85.449072	19-Apr-19	25-Apr-19
314	42.4901646	-85.448943	19-Apr-19	25-Apr-19
335	42.484711	-85.450802	12-Apr-19	25-Apr-19
363	42.4845832	-85.451169	09-May-19	10-May-19
366	42.4878624	-85.448518	09-May-19	10-May-19
385	42.4793761	-85.457742	12-Apr-19	25-Apr-19
386	42.4902613	-85.448932	19-Apr-19	25-Apr-19
391	42.4877635	-85.448291	10-May-19	10-May-19

Table S2.1 Rhizome collection details from mirid-feeding experiment. Putative genotype identification number is arbitrary and based on the physical location where rhizomes were collected, not with molecular techniques.

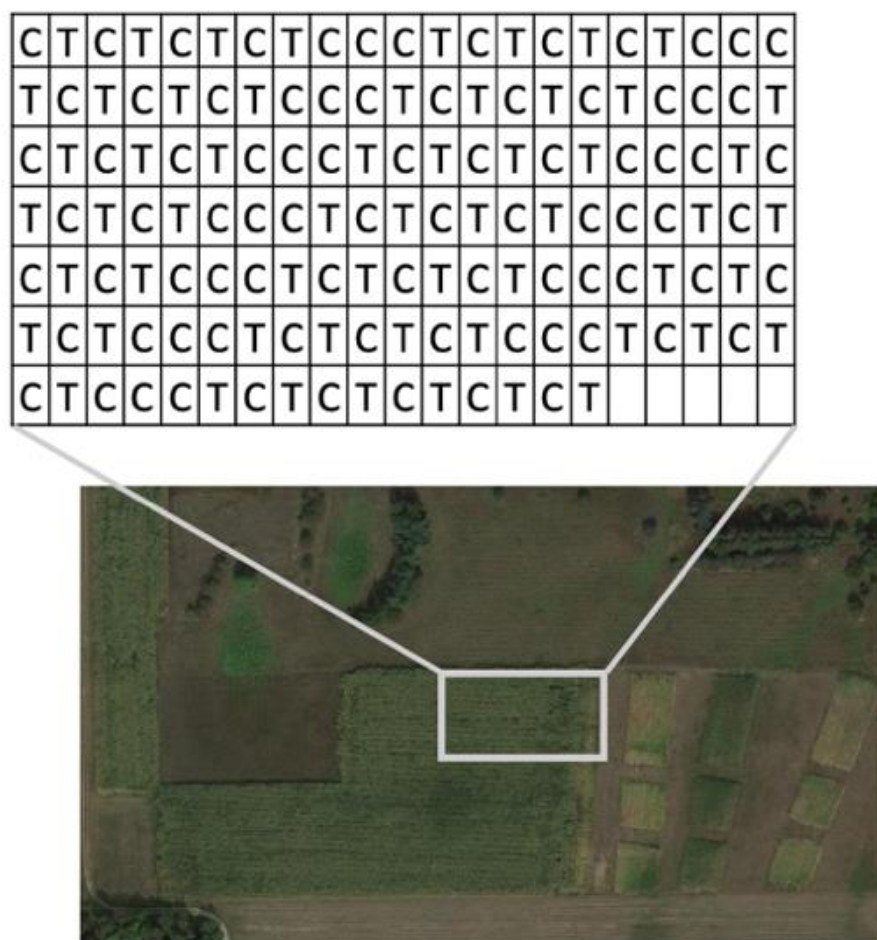


Figure S2.1 Map of mirid-feeding experiment. ‘C’ stands for ‘control’ plants that did not receive mirid-feeding. ‘T’ stands for ‘treatment’ plants, which received mirid feeding.

	Measurement Round 1 (17- 19-Jul-2019)	Measurement Round 2 (12- 14-Aug-2019)	Measurement Round 3 (31- Aug-2019 & 2-Sep-2019)	Measurement Round 4 (21- 22-Sep-2019)
Control	84	84	84	84
Treatment	58	58	58	58

Table S2.2 Sample sizes from mirid-feeding experiment by response variable measurement round.

29-32 ramets with various JA
spray timing, randomly assigned



Figure S2.2 Map from JA spray experiment in Lux Arbor Forest Reserve, Kellogg Biological Station, Michigan, USA.

	Round 1 (18-Jul-2020)	Round 2 (31-Jul-2020)	Round 3 (14-15-Aug-2020)	Round 4 (30-Aug-2020)	Round 5 (09-Sep-2020)
Control (No JA)	126	111	103	93	93
Early-only	34	27	27	19	19
Middle-only			23	15	15
Early + Middle			7	7	7
Late-only				10	10
Early + Late				8	8
Middle + Late				8	8

Table S2.3 Sample sizes JA spray experiment by response variable measurement round and spray timing. Blank merged cells are left intentionally blank because that spray timing was not applicable at that measurement round.

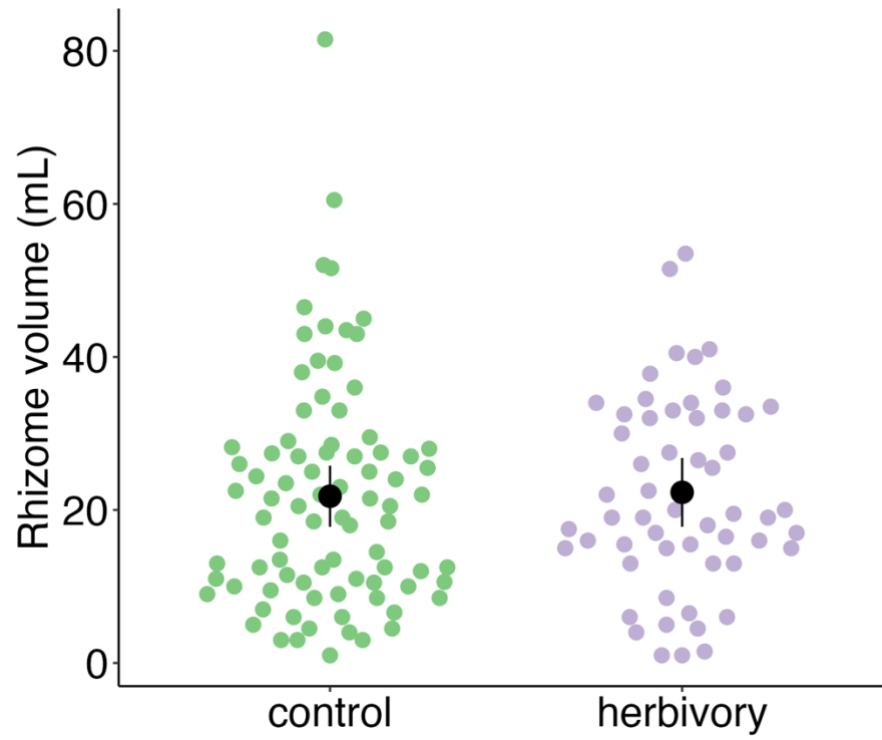


Figure S2.3 Rhizome volume from mirid-feeding experiment.

Response variable	Marginal R^2
Chewing herbivory	31.9%
Pathogen damage	51.0%
Plant height	26.5%

Table S2.4 Variance explained (marginal R^2) for each model in the mirid-feeding experiment.

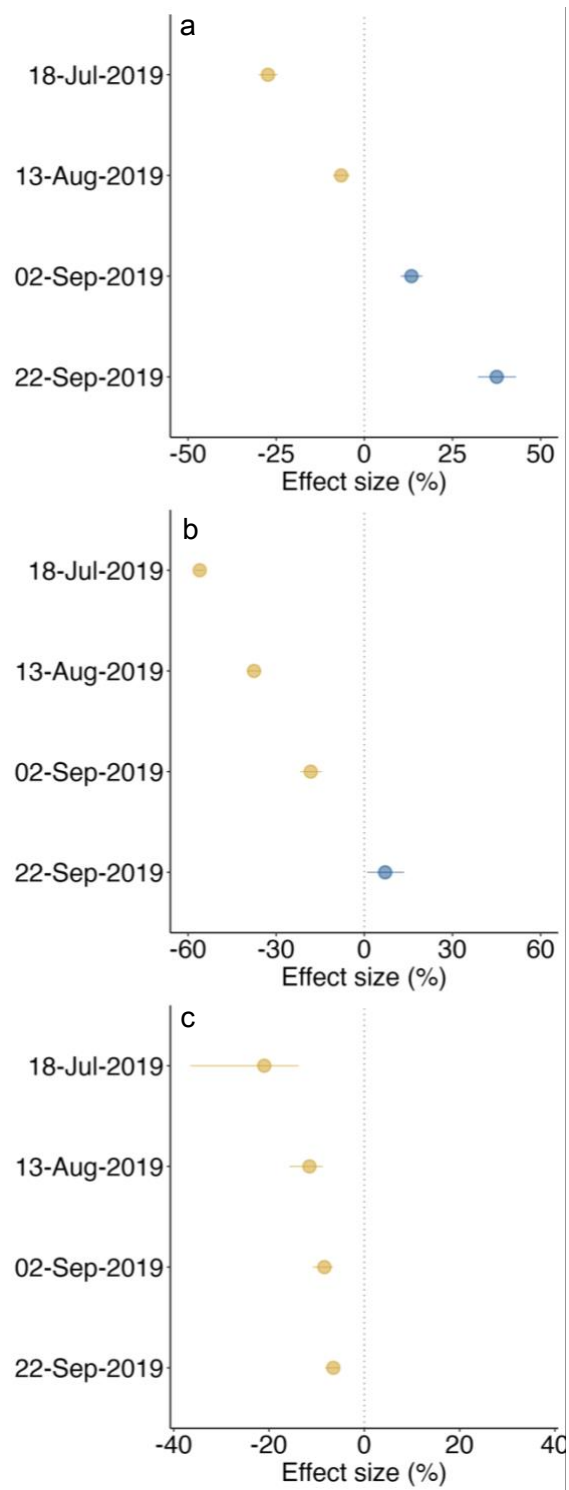


Figure S2.4 Plotted effect sizes for each mirid-feeding experiment on (a) chewing herbivory, (b) pathogen damage, and (c) plant height. Shown are the model predicted means (points) and 95% CI (lines). Yellow indicates significantly negative effects on the measured parameters, and blue indicates significantly positive effects.

(a) *Chewing herbivory*

<i>Date</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
18-Jul-2019	-27.3	-30.0	-24.5	Significantly Negative
13-Aug-2019	-6.5	-8.9	-4.3	Significantly Negative
02-Sep-2019	13.4	10.3	16.5	Significantly Positive
22-Sep-2019	37.6	32.3	43.0	Significantly Positive

(b) *Pathogen damage*

<i>Date</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
18-Jul-2019	-56.0	-57.6	-54.4	Significantly Negative
13-Aug-2019	-37.3	-39.8	-35.2	Significantly Negative
02-Sep-2019	-18.2	-21.9	-14.4	Significantly Negative
22-Sep-2019	7.1	1.0	13.6	Significantly Positive

(c) *Plant height*

<i>Date</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
18-Jul-2019	-21.0	-36.5	-13.8	Significantly Negative
13-Aug-2019	-11.5	-15.7	-8.7	Significantly Negative
02-Sep-2019	-8.4	-10.8	-6.6	Significantly Negative
22-Sep-2019	-6.5	-8.2	-5.1	Significantly Negative

(d) *Rhizome volume*

<i>Date</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
13-Dec-2019	0.5	0.0	1.0	Not significant

Table S2.5 Effect sizes for the mirid-feeding experiment in tabular form.

Response variable	Observation Timing	Marginal R^2
Chewing herbivory	Early	0.4%
Chewing herbivory	Mid	1.3%
Chewing herbivory	Late	2.9%
Pathogen damage	Early	7.1%
Pathogen damage	Mid	3.0%
Pathogen damage	Late	3.3%
Plant height	Early	11.6%
Plant height	Mid	0.7%
Plant height	Late	1.6%

Table S2.6 Variance explained (marginal and conditional R^2) for each model.

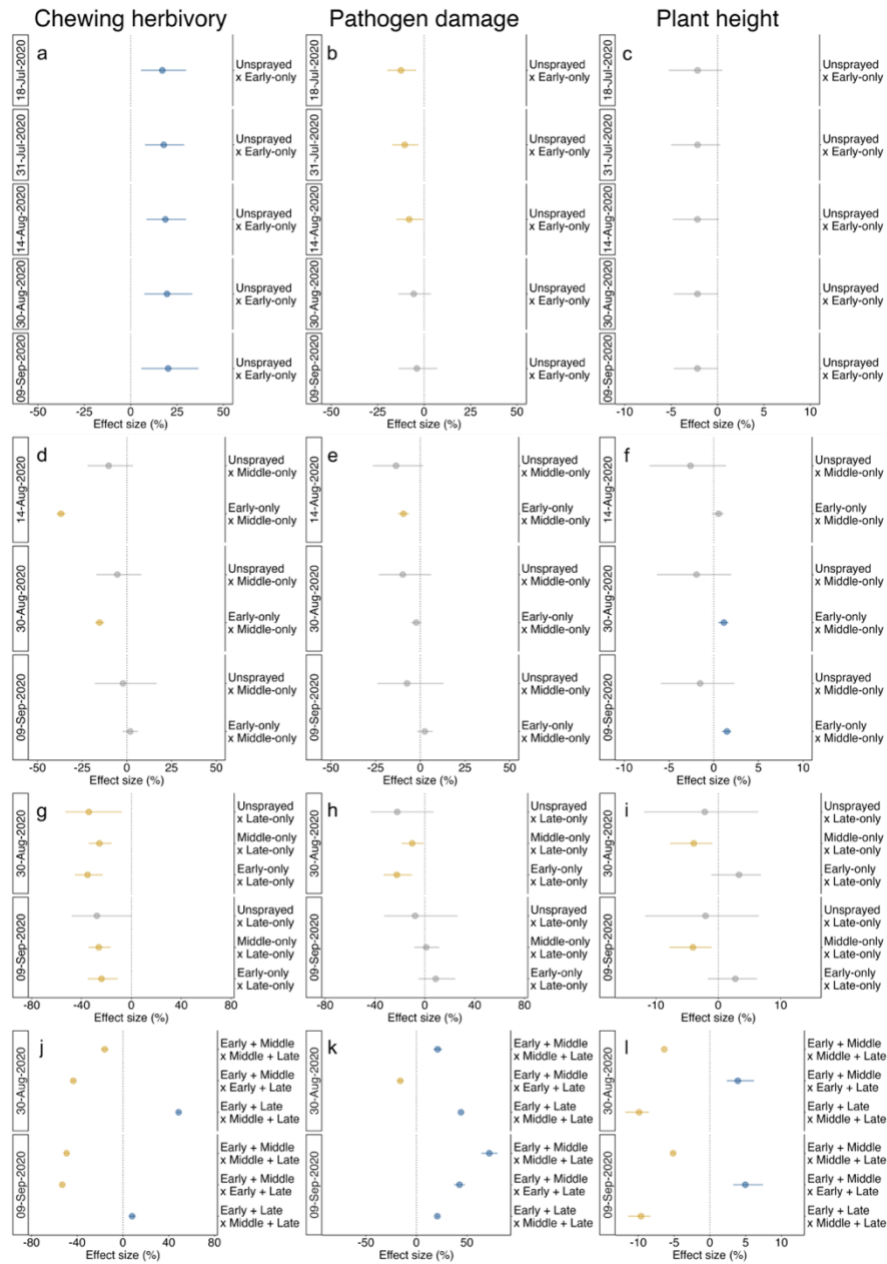


Figure S2.5 Effect sizes with 95% CI between unsprayed plants and plants with JA sprays, treated at various timings faceted by applicable observation date, for a-c) early spray timing, d-f) middle spray timing, g-i) late spray timing, and j-l) double spray timing. Effect sizes are presented as the mean predicted values (dot) and 95% CI (line). Yellow indicates significantly negative effects on the measured parameters, and blue indicates significantly positive effects. Gray dots and lines represent not significant results.

(a) *Chewing herbivory – early*

Date	Comparison	Mean Effect (%)	2.5% CI (%)	97.5% CI (%)	Significance
18-Jul-2020	<i>Unsprayed</i> x <i>Early-only</i>	17.0	5.6	29.8	Significantly Positive
31-Jul-2020	<i>Unsprayed</i> x <i>Early-only</i>	17.8	7.8	28.8	Significantly Positive
14-Aug-2020	<i>Unsprayed</i> x <i>Early-only</i>	18.7	8.5	29.8	Significantly Positive
30-Aug-2020	<i>Unsprayed</i> x <i>Early-only</i>	19.6	7.4	33.1	Significantly Positive
09-Sep-2020	<i>Unsprayed</i> x <i>Early-only</i>	20.2	5.8	36.5	Significantly Positive

(b) *Chewing herbivory – mid*

Date	Comparison	Mean Effect (%)	2.5% CI (%)	97.5% CI (%)	Significance
14-Aug-2020	<i>Unsprayed</i> x <i>Middle-only</i>	-10.2	-22.0	3.4	Not Significant
14-Aug-2020	<i>Unsprayed</i> x <i>Early + Middle</i>	76.7	20.0	160.0	Significantly Positive
14-Aug-2020	<i>Early-only</i> x <i>Middle-only</i>	-36.9	-39.4	-34.3	Significantly Negative
14-Aug-2020	<i>Early-only</i> x <i>Early + Middle</i>	24.1	-6.7	65.2	Not Significant
14-Aug-2020	<i>Middle-only</i> x <i>Early + Middle</i>	96.7	53.9	151.5	Significantly Positive
30-Aug-2020	<i>Unsprayed</i> x <i>Middle-only</i>	-5.3	-17.0	8.0	Not Significant
30-Aug-2020	<i>Unsprayed</i> x <i>Early + Middle</i>	26.9	-8.2	75.3	Not Significant
30-Aug-2020	<i>Early-only</i> x <i>Middle-only</i>	-15.3	-17.7	-12.7	Significantly Negative
30-Aug-2020	<i>Early-only</i> x <i>Early + Middle</i>	13.6	-9.0	41.6	Not Significant
30-Aug-2020	<i>Middle-only</i> x <i>Early + Middle</i>	34.0	10.7	62.3	Significantly Positive
09-Sep-2020	<i>Unsprayed</i> x <i>Middle-only</i>	-2.2	-17.9	16.6	Not Significant
09-Sep-2020	<i>Unsprayed</i> x <i>Early + Middle</i>	3.2	-28.5	48.9	Not Significant
09-Sep-2020	<i>Early-only</i> x <i>Middle-only</i>	1.9	-2.3	6.2	Not Significant
09-Sep-2020	<i>Early-only</i> x <i>Early + Middle</i>	7.4	-14.9	35.7	Not Significant
09-Sep-2020	<i>Middle-only</i> x <i>Early + Middle</i>	5.5	-12.9	27.7	Not Significant

Table S2.7 Effect sizes for the JA-spray experiment in tabular form.

Table S2.7 (cont'd)(c) *Chewing herbivory – late*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
30-Aug-2020	<i>Unsprayed x Late-only</i>	-34.2	-53.0	-7.7	Significantly Negative
30-Aug-2020	<i>Unsprayed x Early + Late</i>	-34.1	-55.3	-2.9	Significantly Negative
30-Aug-2020	<i>Unsprayed x Middle + Late</i>	-2.1	-33.1	43.3	Not Significant
30-Aug-2020	<i>Early-only x Late-only</i>	-35.2	-45.4	-23.1	Significantly Negative
30-Aug-2020	<i>Early-only x Early + Late</i>	-35.1	-48.0	-19.1	Significantly Negative
30-Aug-2020	<i>Early-only x Middle + Late</i>	-3.6	-22.2	19.4	Not Significant
30-Aug-2020	<i>Middle-only x Late-only</i>	-25.7	-34.3	-15.9	Significantly Negative
30-Aug-2020	<i>Middle-only x Early + Late</i>	-25.6	-37.5	-11.5	Significantly Negative
30-Aug-2020	<i>Middle-only x Middle + Late</i>	10.5	-6.5	30.6	Not Significant
30-Aug-2020	<i>Early + Middle x Early + Late</i>	-43.4	-41.7	-45.1	Significantly Negative
30-Aug-2020	<i>Early + Middle x Middle + Late</i>	-16.0	-12.8	-19.0	Significantly Negative
30-Aug-2020	<i>Late-only x Middle + Late</i>	48.7	42.3	55.2	Significantly Positive
30-Aug-2020	<i>Late-only x Early + Late</i>	0.05	-4.8	5.2	Not Significant
30-Aug-2020	<i>Late-only x Early + Middle</i>	76.9	63.2	91.7	Significantly Positive
30-Aug-2020	<i>Early + Late x Middle + Late</i>	48.6	49.6	47.6	Significantly Positive
09-Sep-2020	<i>Unsprayed x Late-only</i>	-27.7	-47.9	0.2	Not Significant
09-Sep-2020	<i>Unsprayed x Early + Late</i>	-48.8	-65.6	-23.9	Significantly Negative
09-Sep-2020	<i>Unsprayed x Middle + Late</i>	-44.7	-62.0	-19.6	Significantly Negative
09-Sep-2020	<i>Early-only x Late-only</i>	-24.0	-35.0	-11.0	Significantly Negative
09-Sep-2020	<i>Early-only x Early + Late</i>	-46.1	-57.1	-32.4	Significantly Negative
09-Sep-2020	<i>Early-only x Middle + Late</i>	-41.8	-52.6	-28.6	Significantly Negative

Table S2.7 (cont'd)

09-Sep-2020	<i>Middle-only x Late-only</i>	-26.2	-34.4	-16.9	Significantly Negative
09-Sep-2020	<i>Middle-only x Early + Late</i>	-47.7	-56.6	-36.9	Significantly Negative
09-Sep-2020	<i>Middle-only x Middle + Late</i>	-43.5	-52.1	-33.3	Significantly Negative
09-Sep-2020	<i>Early + Middle x Early + Late</i>	-53.0	-51.8	-54.2	Significantly Negative
09-Sep-2020	<i>Early + Middle x Middle + Late</i>	-49.3	-46.8	-51.6	Significantly Negative
09-Sep-2020	<i>Late-only x Middle + Late</i>	-23.5	-27.0	-19.8	Significantly Negative
09-Sep-2020	<i>Late-only x Early + Late</i>	-29.1	-33.9	-24.0	Significantly Negative
09-Sep-2020	<i>Late-only x Early + Middle</i>	50.8	37.1	65.9	Significantly Positive
09-Sep-2020	<i>Early + Late x Middle + Late</i>	8.0	10.5	5.6	Significantly Positive

(d) *Pathogen damage – early*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
18-Jul-2020	<i>Unsprayed x Early-only</i>	-12.4	-19.9	-4.3	Significantly Negative
31-Jul-2020	<i>Unsprayed x Early-only</i>	-10.4	-17.2	-3.1	Significantly Negative
14-Aug-2020	<i>Unsprayed x Early-only</i>	-8.0	-15.0	-0.5	Significantly Negative
30-Aug-2020	<i>Unsprayed x Early-only</i>	-5.6	-13.9	3.6	Not Significant
09-Sep-2020	<i>Unsprayed x Early-only</i>	-3.9	-13.7	7.0	Not Significant

Table S2.7 (cont'd)(e) *Pathogen damage – mid*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
14-Aug-2020	<i>Unsprayed x Middle-only</i>	-13.6	-26.5	1.6	Not Significant
14-Aug-2020	<i>Unsprayed x Early + Middle</i>	-2.1	-34.4	46.1	Not Significant
14-Aug-2020	<i>Early-only x Middle-only</i>	-9.4	-12.4	-6.4	Significantly Negative
14-Aug-2020	<i>Early-only x Early + Middle</i>	2.5	-21.9	34.6	Not Significant
14-Aug-2020	<i>Middle-only x Early + Middle</i>	13.2	-10.9	43.8	Not Significant
30-Aug-2020	<i>Unsprayed x Middle-only</i>	-9.8	-23.3	6.0	Not Significant
30-Aug-2020	<i>Unsprayed x Early + Middle</i>	-30.0	-51.1	0.4	Not Significant
30-Aug-2020	<i>Early-only x Middle-only</i>	-2.3	-5.1	0.6	Not Significant
30-Aug-2020	<i>Early-only x Early + Middle</i>	-24.1	-39.5	-4.7	Significantly Negative
30-Aug-2020	<i>Middle-only x Early + Middle</i>	-22.3	-36.3	-5.3	Significantly Negative
09-Sep-2020	<i>Unsprayed x Middle-only</i>	-7.4	-24.1	12.9	Not Significant
09-Sep-2020	<i>Unsprayed x Early + Middle</i>	-43.2	-61.4	-16.4	Significantly Negative
09-Sep-2020	<i>Early-only x Middle-only</i>	2.5	-1.7	6.9	Not Significant
09-Sep-2020	<i>Early-only x Early + Middle</i>	-37.1	-50.0	-20.8	Significantly Negative
09-Sep-2020	<i>Middle-only x Early + Middle</i>	-38.6	-49.2	-25.9	Significantly Negative

Table S2.7 (cont'd)(f) *Pathogen damage – late*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
30-Aug-2020	<i>Unsprayed x Late-only</i>	-22.2	-43.4	6.9	Not Significant
30-Aug-2020	<i>Unsprayed x Early + Late</i>	-30.6	-51.9	0.1	Not Significant
30-Aug-2020	<i>Unsprayed x Middle + Late</i>	-0.1	-30.9	44.6	Not Significant
30-Aug-2020	<i>Early-only x Late-only</i>	-22.5	-33.0	-10.3	Significantly Negative
30-Aug-2020	<i>Early-only x Early + Late</i>	-30.9	-43.1	-16.0	Significantly Negative
30-Aug-2020	<i>Early-only x Middle + Late</i>	-0.4	-18.3	21.3	Not Significant
30-Aug-2020	<i>Middle-only x Late-only</i>	-10.1	-18.5	-1.0	Significantly Negative
30-Aug-2020	<i>Middle-only x Early + Late</i>	-19.9	-30.8	-7.3	Significantly Negative
30-Aug-2020	<i>Middle-only x Middle + Late</i>	15.4	-0.6	34.0	Not Significant
30-Aug-2020	<i>Early + Middle x Early + Late</i>	-16.0	-12.5	-19.1	Significantly Negative
30-Aug-2020	<i>Early + Middle x Middle + Late</i>	21.0	25.3	16.9	Significantly Positive
30-Aug-2020	<i>Late-only x Middle + Late</i>	28.4	21.9	35.3	Significantly Positive
30-Aug-2020	<i>Late-only x Early + Late</i>	-10.9	-15.1	-6.4	Significantly Negative
30-Aug-2020	<i>Late-only x Early + Middle</i>	6.1	-2.7	15.7	Not Significant
30-Aug-2020	<i>Early + Late x Middle + Late</i>	44.1	43.7	44.5	Significantly Positive
09-Sep-2020	<i>Unsprayed x Late-only</i>	-7.8	-32.6	26.1	Not Significant
09-Sep-2020	<i>Unsprayed x Early + Late</i>	-25.8	-48.8	7.5	Not Significant
09-Sep-2020	<i>Unsprayed x Middle + Late</i>	-10.5	-37.6	28.5	Not Significant
09-Sep-2020	<i>Early-only x Late-only</i>	8.7	-5.1	24.4	Not Significant
09-Sep-2020	<i>Early-only x Early + Late</i>	-12.5	-27.8	6.0	Not Significant
09-Sep-2020	<i>Early-only x Middle + Late</i>	5.6	-12.1	26.8	Not Significant

Table S2.7 (cont'd)

09-Sep-2020	<i>Middle-only x Late-only</i>	1.1	-8.3	11.5	Not Significant
09-Sep-2020	<i>Middle-only x Early + Late</i>	-18.6	-30.3	-5.0	Significantly Negative
09-Sep-2020	<i>Middle-only x Middle + Late</i>	-1.8	-15.1	13.6	Not Significant
09-Sep-2020	<i>Early + Middle x Early + Late</i>	42.6	48.1	37.4	Significantly Positive
09-Sep-2020	<i>Early + Middle x Middle + Late</i>	72.1	80.3	64.3	Significantly Positive
09-Sep-2020	<i>Late-only x Middle + Late</i>	-2.9	-7.4	1.9	Not Significant
09-Sep-2020	<i>Late-only x Early + Late</i>	-19.5	-24.0	-14.8	Significantly Negative
09-Sep-2020	<i>Late-only x Early + Middle</i>	-43.6	-48.6	-38.0	Significantly Negative
09-Sep-2020	<i>Early + Late x Middle + Late</i>	20.7	21.8	19.6	Significantly Positive

(g) *Plant height – early*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
18-Jul-2020	<i>Unsprayed x Early-only</i>	-2.2	-5.3	0.5	Not Significant
31-Jul-2020	<i>Unsprayed x Early-only</i>	-2.2	-5.0	0.4	Not Significant
14-Aug-2020	<i>Unsprayed x Early-only</i>	-2.2	-4.8	0.1	Not Significant
30-Aug-2020	<i>Unsprayed x Early-only</i>	-2.2	-4.71	0.1	Not Significant
09-Sep-2020	<i>Unsprayed x Early-only</i>	-2.2	-4.69	0.0	Not Significant

Table S2.7 (cont'd)(h) *Plant height – mid*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
14-Aug-2020	<i>Unsprayed x Middle-only</i>	-2.6	-7.2	1.4	Not Significant
14-Aug-2020	<i>Unsprayed x Early + Middle</i>	-4.8	-16.9	5.7	Not Significant
14-Aug-2020	<i>Early-only x Middle-only</i>	0.5	-0.2	1.1	Not Significant
14-Aug-2020	<i>Early-only x Early + Middle</i>	-1.8	-10.7	5.4	Not Significant
14-Aug-2020	<i>Middle-only x Early + Middle</i>	-2.3	-10.5	4.3	Not Significant
30-Aug-2020	<i>Unsprayed x Middle-only</i>	-1.9	-6.4	1.9	Not Significant
30-Aug-2020	<i>Unsprayed x Early + Middle</i>	-3.9	-15.6	6.3	Not Significant
30-Aug-2020	<i>Early-only x Middle-only</i>	1.1	0.5	1.6	Significantly Positive
30-Aug-2020	<i>Early-only x Early + Middle</i>	-0.9	-9.4	6.0	Not Significant
30-Aug-2020	<i>Middle-only x Early + Middle</i>	-2.0	-9.9	4.3	Not Significant
09-Sep-2020	<i>Unsprayed x Middle-only</i>	-1.5	-5.9	2.3	Not Significant
09-Sep-2020	<i>Unsprayed x Early + Middle</i>	-3.3	-14.9	6.8	Not Significant
09-Sep-2020	<i>Early-only x Middle-only</i>	1.5	0.9	1.9	Significantly Positive
09-Sep-2020	<i>Early-only x Early + Middle</i>	-0.4	-8.7	6.4	Not Significant
09-Sep-2020	<i>Middle-only x Early + Middle</i>	-1.8	-9.5	4.4	Not Significant

Table S2.7 (cont'd)(i) *Plant height – late*

<i>Date</i>	<i>Comparison</i>	<i>Mean Effect (%)</i>	<i>2.5% CI (%)</i>	<i>97.5% CI (%)</i>	<i>Significance</i>
30-Aug-2020	<i>Unsprayed x Late-only</i>	-2.2	-11.9	6.4	Not Significant
30-Aug-2020	<i>Unsprayed x Early + Late</i>	-0.6	-11.5	9.1	Not Significant
30-Aug-2020	<i>Unsprayed x Middle + Late</i>	-10.4	-22.0	-0.2	Significantly Negative
30-Aug-2020	<i>Early-only x Late-only</i>	3.3	-1.1	6.8	Not Significant
30-Aug-2020	<i>Early-only x Early + Late</i>	5.0	-0.7	9.5	Not Significant
30-Aug-2020	<i>Early-only x Middle + Late</i>	-5.4	-12.4	0.2	Not Significant
30-Aug-2020	<i>Middle-only x Late-only</i>	-3.9	-7.7	-1.0	Significantly Negative
30-Aug-2020	<i>Middle-only x Early + Late</i>	-2.4	-7.3	1.5	Not Significant
30-Aug-2020	<i>Middle-only x Middle + Late</i>	-12.0	-18.3	-7.1	Significantly Negative
30-Aug-2020	<i>Early + Middle x Early + Late</i>	3.9	6.2	2.4	Significantly Positive
30-Aug-2020	<i>Early + Middle x Middle + Late</i>	-6.4	-6.4	-6.4	Significantly Negative
30-Aug-2020	<i>Late-only x Middle + Late</i>	-8.4	-11.4	-6.2	Significantly Negative
30-Aug-2020	<i>Late-only x Early + Late</i>	1.6	0.4	2.5	Significantly Positive
30-Aug-2020	<i>Late-only x Early + Middle</i>	-2.2	-5.4	0.1	Not Significant
30-Aug-2020	<i>Early + Late x Middle + Late</i>	-9.9	-11.8	-8.5	Significantly Negative
09-Sep-2020	<i>Unsprayed x Late-only</i>	-2.1	-11.7	6.5	Not Significant
09-Sep-2020	<i>Unsprayed x Early + Late</i>	-0.1	-11.1	9.5	Not Significant
09-Sep-2020	<i>Unsprayed x Middle + Late</i>	-9.7	-21.2	0.4	Not Significant
09-Sep-2020	<i>Early-only x Late-only</i>	2.7	-1.7	6.2	Not Significant
09-Sep-2020	<i>Early-only x Early + Late</i>	4.7	-0.9	9.2	Not Significant
09-Sep-2020	<i>Early-only x Middle + Late</i>	-5.3	-12.3	0.1	Not Significant
09-Sep-2020	<i>Middle-only x Late-only</i>	-4.1	-7.9	-1.1	Significantly Negative
09-Sep-2020	<i>Middle-only x Early + Late</i>	-2.2	-7.2	1.7	Not Significant
09-Sep-2020	<i>Middle-only x Middle + Late</i>	-11.6	-17.8	-6.8	Significantly Negative
09-Sep-2020	<i>Early + Middle x Early + Late</i>	5.0	7.4	3.3	Significantly Positive
09-Sep-2020	<i>Early + Middle x Middle + Late</i>	-5.1	-4.8	-5.4	Significantly Negative

Table S2.7 (cont'd)

09-Sep-2020	<i>Late-only x Middle + Late</i>	-7.9	-10.8	-5.7	Significantly Negative
09-Sep-2020	<i>Late-only x Early + Late</i>	2.0	0.8	2.8	Significantly Positive
09-Sep-2020	<i>Late-only x Early + Middle</i>	-2.9	-6.2	-0.4	Significantly Negative
09-Sep-2020	<i>Early + Late x Middle + Late</i>	-9.6	-11.4	-8.3	Significantly Negative

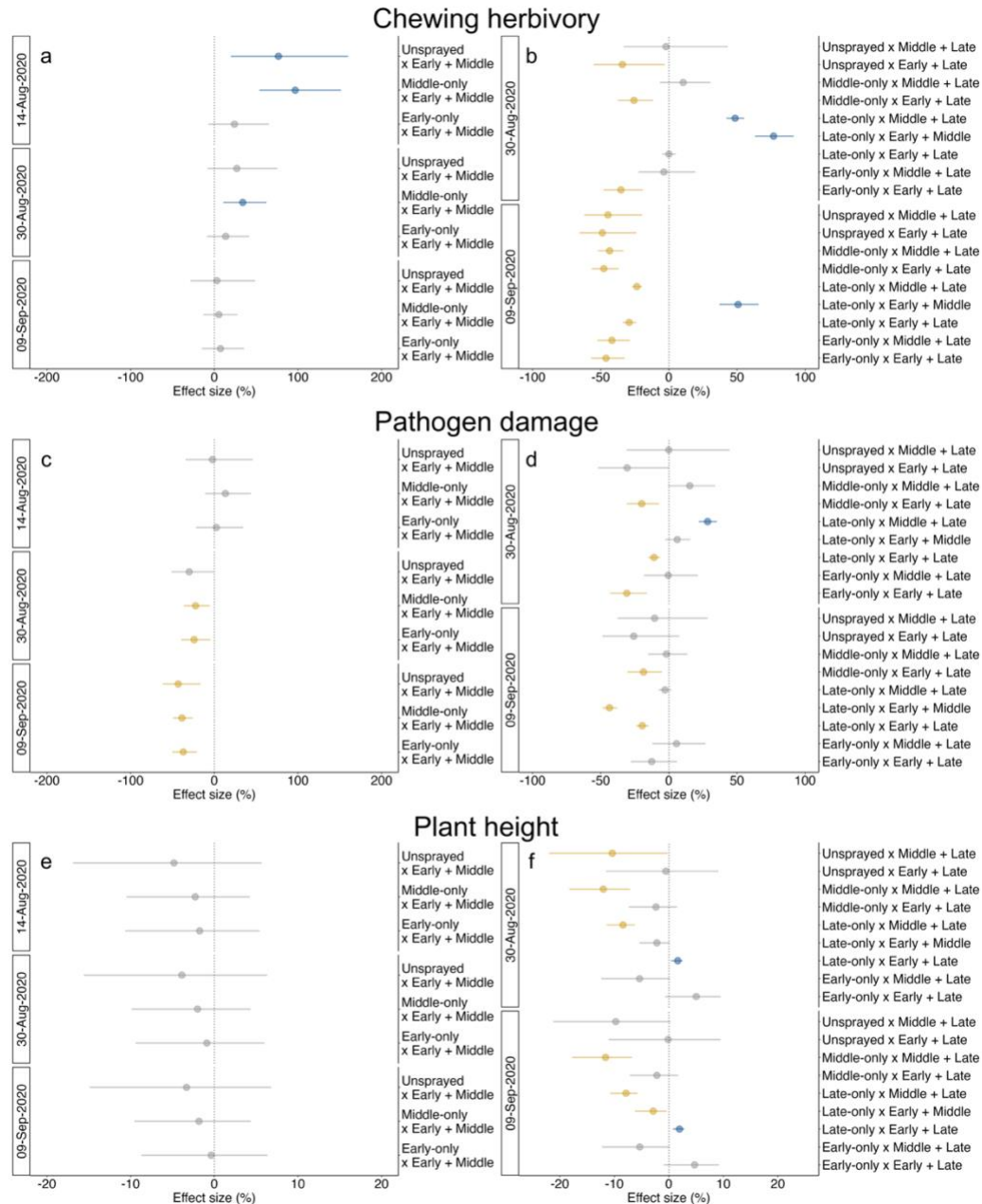


Figure S2.6 Effect sizes with 95% CI between unsprayed plants and plants with JA sprays, treated at various frequencies faceted by applicable observation date, for a, c, e) middle spray timing and b, d, f) late spray timing. Effect sizes are presented as the mean predicted values (dot) and 95% CI (line). Yellow indicates significantly negative effects on the measured parameters, and blue indicates significantly positive effects. Gray dots and lines represent not significant results.