Medicines, Metabolites, and Pigments in Caryophyllales and Beyond

A Dissertation

SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA

BY

Alexandra Hazel Crum

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Ya Yang

2024

Acknowledgements

I have many people to thank for their support over the last 5 years. First, I must thank Dr. Ya Yang for her support. I've learned so much about the scientific process and how to be a scientist from her. Without her, this project would never have developed past a vague idea about medicinal plants and metabolism. She's helped me figure out how to be a woman in science, and how to manage my time while still supporting my many extracurriculars that helped me figure out where in academia I wanted to fit. I am forever grateful for her investment in me.

I also want to thank my committee members, Dr. Adrian Hegeman, Dr. Clay Carter, and Dr. Lisa Philander, for their guidance throughout this process. They have spent hours discussing research plans and results with me, and have invested in my professional development in and out of the lab. In particular, my ethnobotanical work owes much to Lisa, and my metabolomics survey of Caryophyllales would not exist without Adrian's feedback, advice, time, and sharing of his lab.

Past and present members of the Yang lab have contributed much to this project and my time in Plant & Microbial Biology. Dr. Rebekah Mohn and Dr. Diego Morales-Briones, who offered their time and support from before I had even applied to their departures. Zack Radford was instrumental in developing and optimizing plant germination and stress experiments, helping identify medicinal plants, and in keeping the lab functioning in his time here. Aaron Lee has contributed his time and data analysis expertise to our massive *Silene* and *Beta* transcriptome database, and along with Ya, Zack, and Dr. Brett Fredericksen braved BioSci at its spookiest to sample the midnight timepoints of the *Beta* and *Silene* stress experiments. Visiting lab member Dr. Nan Lin gave support, advice, and lots of delicious food.

Undergraduate members of the lab Nicholas De La Rosa, Kyle Duval, Erin Boehme, and Lauren Vander Esch have given feedback at lab meetings and on talks. Annika Smeenk and Addison assisted with both plant care and plant stress. Annika, Sophie Naylor, and Adiel Andino-Acevedo spent many hours working on PCR and identification of Hmong herbs. Addison also took on a project, that while not included here, is something I had wanted to explore for a long time, and I am grateful for all of the work he has put in, and his ability to take ownership of a project as an undergraduate that would be daunting for many at a master's level.

Leticia Magpali, while participating in lab meetings, was very helpful in suggesting ethnobotanical databases to search outside the U.S. for compiling medicinal uses of Caryophyllales species. Dr. Luke Busta assisted in data collection for the same project with contribution of an R script for querying PubMed. Angie Ricono similarly provided help with analyzing spectral data in R.

I am also grateful to Natalie Hoidal for involving me in a "quick" project identifying Hmong herbs, which has become a chapter here, is becoming a book thanks to Natalie's persistence, and is one of the most rewarding things I have ever been a part of. Thank you to Lindsey Miller for photographing all of the medicinal plants, the process of collecting them, and on one weird day, me pretending to do PCR. Most especially, thank you to Zongxee, May, and the rest of the Lee family for stewarding these herbs and giving your knowledge to the world. I am thrilled I was able to contribute in a small way.

Thank you to Dr. Nox Makunga, Dr. Andrea Berardi, Dr. John Cushman, and the Desert Botanical Garden for sending plant material used in stress experiments. Andrea also contributed *Silene* expertise and genes. Thank you also to Diego Paredes Burneo and Dr. Luis Santiago-

Rosario for finding *Mollugo verticillata* seedlings in Louisiana. Diego's help in collecting them during my less-then-48 hour trip to Baton Rouge was invaluable.

Dr. Kate Freund Saxhaug's time and expertise as I learned how to work the LC-MS and analyze the data from it was invaluable, and I am thankful for her patience and positivity. Dr. Hiroshi Maeda and Madelyn Schaut also ran metabolomics analyses of the *Beta* and *Silene* time course data and have been extremely helpful in discussing the results.

I am grateful to Dr. George Weiblen, for his investment in my development as an instructor and botanist. I learned so much teaching Minnesota Flora during the pandemic that I will take with me into all future classes, and the U.S. Army Corps Biodiversity Survey was the perfect introduction to grad school. Thank you also to Tim Whitfeld, for the time he spent training me in the herbarium my first year.

Thank you to all of the collaborators on the Caryophyllales project I have yet to mention. This includes Dr. Sam Brockington, Dr. Boas Pucker, Dr. Nat Walker-Hale, Dr. Beth Moore, Dr. Hester Sheehan, and more. They have shared feedback, data, and seeds.

Over the course of my Ph.D, I have had the pleasure of working at the CBS Conservatory. Thank you to Jared Rubinstein, Adam Wegren, Betsy Custis, and the many conservatory assistants that have passed through. I am forever grateful for the opportunity to work with such an incredible group of people. Their support, and being able to work with the living collection, has meant so much, and I am a better botanist for it. Also, through the Conservatory, I met Dr. Jake Grossman, who I thank for his career advice and support in finding a position at my ideal institution.

I wouldn't be writing this dissertation today without the support of my friends and family. The moral support of my friends has gotten me through a lot and helped make the Twin Cities

feel like home. My family is the reason I became interested in science. My parents, Kim and Craig Giermann, let me look through the microscopes in their classrooms as a kid, and used to pay me a dime for every plant I could identify in the backyard. They knew long before I did that I wanted to be a botanist. I hope that as Jake and Jonah also become doctors (of a different kind) I am as much of a support to them as they have been to me. Thank you also to my grandfather, Yale Pulling, for giving many of my stressed plants a happy retirement. My tortoises, Starbuck and Helo, and dog Jamie have also been critical supports in writing this dissertation.

Finally, I thank the many, many indigenous cultures and communities that have shared their knowledge, without which, much of this work would not have been possible. My research incorporated knowledge held by communities across the globe. However, I especially owe thanks to the Hmong people. Additionally, much of the research presented here was conducted on the traditional homelands of the Dakota people. It is important to acknowledge the peoples on whose land we live, learn, and work as we seek to improve and strengthen our relations with our tribal nations.

This work has been supported by the National Science Foundation (NSFDEB-NERC 1939226 and NSF-DBI 2021898), SOAR-REEU, and the Alexander P. Anderson and Lydia Anderson Grant.

Dedication

Dedicated firstly to my parents. Thank you for inspiring a sense of wonder and making everything a science experiment.

Dedicated also to my grandma. Thank you for showing me the edible violets, kicking off a lifelong obsession.

Table of Contents

INTRODUCTION

Much of the world still relies on traditional systems of primarily plant-based medicine (Hamilton, 2004; Gaoue et al. 2021). Even those that rely on modern pharmaceuticals benefit from medicines derived from plants, such as the cancer drugs paclitaxel, vinicristine, and vinblastine, the pain medication aspirin, or the anti-malarial artemisinin. However, the forces that drive the evolution of certain plants to be medicinal or to be used over others have not been studied in a comprehensive way. Medicinal plant selection is likely driven by a combination of cultural factors, biogeography, phylogeny, and specialized metabolism (Gaoue et al., 2017; Teixidor-Toneu et al., 2018). By examining medicinal plants at multiple scales, including at a global scale, within a community, and at a genetic and metabolic level, a full picture of how plants become medicinal can be gleaned. In addition, understanding specialized metabolism has great utility outside of human medicine, as the evolution of different specialized metabolic pathways is complex and interconnected, and these compounds are involved with not only an individual plant's life cycle, but also how plants interact with their wider environment.

Specialized metabolism for the plant and people

Plant specialized metabolites are important for plant survival, helping mediate biotic and abiotic stressors, such as providing protection from UV light. These compounds are diverse, and it is estimated the plant kingdom produces over one million unique specialized metabolites (Afendi et al., 2012). Some, like anthocyanin pigment, are phylogenetically widespread, while others, like the glucosinolates of Brassicales, are lineage-specific.

While specialized metabolites are vital adaptations for the plant, they also are important to human health, underpinning both traditional medicine and many of our modern drugs. For example, the 2015 Nobel Prize in Medicine was awarded for the discovery of an anti-malarial

compound isolated from a plant used in Traditional Chinese Medicine (Normile, 2015). Plant species used in traditional medicine across the globe are clustered into certain phylogenetic lineages, with a species tending to be more closely related to plants with similar use than random chance (Saslis-Lagoudakis et al., 2012). The phylogenetic clustering of medicinal plants suggests similar specialized metabolites between closely related species.

Caryophyllales as a system for understanding specialized metabolism and patterns of medicinal use

The order Caryophyllales is notable for its core clade's unique evolution of betalain pigments, a specialized metabolite derived from the amino acid tyrosine. The betalain pigments are mutually exclusive with anthocyanins (Phenylalanine-derived pigments), meaning that betalain-producing species have never been shown to produce anthocyanin, and vice versa (Clement & Mabry, 1996). Betalain likely originated approximately four times in Caryophyllales, with some lineages once thought to be reversals back to anthocyanins now believed to have the ancestral pigment state (Lopez Nieves et al., 2018; Sheehan et al., 2020). Both anthocyanin and betalain help the plant mediate abiotic stress, functioning as antioxidants and providing protection from high light conditions by reducing photons pressuring PSII (Jain and Gould, 2015).

Although Caryophyllales is best known for betalain, the pigment is thought to represent a broader diversification of Tyr-derived specialized metabolites in the core Caryophyllales (Lopez-Nieves et al., 2018). Other Tyr-derived metabolites that accumulate in Caryophyllales include isoquinoline alkaloids and catecholamines (Chen et al., 2003). This diversification appears to have been enabled by the duplication of arogenate dehydrogenase (ADH), the gene responsible for converting the precursor arogenate to Tyr, within the core Caryophyllales prior to the origin

of betalain (Lopez-Nieves et al., 2018). Whereas the canonical version (ADH-β) is feedback-inhibited by Tyr, the Caryophyllales-specific copy (ADH-α) has relaxed sensitivity to Tyr inhibition, resulting in higher concentrations of Tyr required for enzyme inhibition and increased availability of the amino acid for metabolite diversification (Lopez-Nieves et al., 2018).

In addition to its role in Tyr synthesis, arogenate is also a precursor to phenylalanine (Phe), the amino acid from which anthocyanins, flavonoids, and other phenylpropanoid compounds are derived. With the duplication and neofunctionalization of ADH, there is likely less arogenate available for the production of Phe and its metabolites. This competition for arogenate between the Tyr and Phe pathways has not only enabled the evolution of betalain and other Tyr-derived metabolites in the order, but also potentially limited the diversity of Phe-derived metabolites and their abundance in betalain-producing species.

Although the effect of different specialized metabolites on human health is varied, there are a few trends. Phe-derived metabolites, in particular flavonoids, tend to confer cardiovascular benefits and have antioxidant, anti-microbial, and anti-cancer properties (Ververidis et al., 2007). While some Tyr metabolites, such as betalain, also act as antioxidants, other groups, like the catecholamines, can act as human neurotransmitters (Schenk and Maeda, 2018). Tyr-derived isoquinoline alkaloids have a range of neurological effects, including the hallucinations induced by mescaline. Given the different action of Phe and Tyr-derived metabolites, the potential changes in Phe metabolism and diversification of Tyr metabolism are expected to have impacted the ways human use species in the ancestral, betalain, and reversal lineages.

This dissertation seeks to understand how plants become medicinal by investigating at several scales. Additionally, the use of Caryophyllales as a study system in three out of four chapters allows for a related and complementary investigation of specialized metabolite

evolution within a phylogenetic framework. The repeated transitions of dominant metabolism type at a scale that encompasses the globe but isn't prohibitive to study will help inform our understanding of medicinal plants while also contributing to a greater understanding of the order and the incredible diversity and complexity of specialized metabolism in its own right. Chapter 1 approaches medicinal plants from a cultural perspective, by using DNA barcoding to identify plants in the Hmong pharmacopeia in Saint Paul, MN. Chapter 2 introduces Caryophyllales as a study system and explores medicinal plant use at the global scale, using phylogeny and biodiversity data to find human interaction with a plant species to be a better predictor of medicinal use than metabolism type. Chapter 3 uses comparative multi-omics methods to investigate pathways associated with anthocyanin and betalain synthesis in Caryophyllales, and Chapter 4 uses untargeted metabolomics to investigate the broad diversity of specialized metabolism in Caryophyllales.

REFERENCES

Afendi, F.M., T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, et al. 2012. KNApSAck Family Databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53: e1.

Chen J., Y.P. Shi, J.Y. Liu. 2003. Determination of Noradrenaline and Dopamine in Chinese Herbal Extracts from *Portulaca oleracea* L. by High Performance Liquid Chromatography. *Journal of Chromatography* 1003: 127-132.

Clement, J.S., and T.J. Mabry. 1996. Pigment evolution in the Caryophyllales: a systematic overview. *Bot. Acta* 109: 360-367.

Gaoue, O.G., M.A. Coe, M. Bond, G. Hart, B. C. Seyler, H. McMillen. 2017. Theories and major hypotheses in ethnobotany. *Economic Botany* 71: 269-287.

Gaoue, O. G., K. Yessoufou, L. Mankga, and F. Vodouhe. 2021. Phylogeny reveals non-random medicinal plant organ selection by local people in Benin. *Plants People Planet* 3: 710–720.

Hamilton, A. C. 2004. Medicinal plants, conservation and livelihoods. *Biodiversity and Conservation* 13: 1477–1517.

Jain, G. and K.S. Gould. 2015. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany* 119: 48-53.

Lopez-Nieves, S., Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytologist 217: 896–908.

Normile D. 2015. Nobel for antimalarial drug highlights East-West divide. *Science* 350: 265.

Saslis-Lagoudakis C.H., V. Savolainen, E.M. Williamson, F. Forest, S.J. Wagstaff, S.R. Baral, M.F. Watson, C.A. Pendry, J.A. Hawkins. 2012. Phylogenies reveal predictive power of traditional medicine in bioprospecting. *PNAS* 109: 15835-15840.

Sheehan, H., T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo. W.C. Yim, et al. 2020. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent specialization to betalain pigmentation in Caryophyllales. *New Phytologist* 227: 914-929.

Schenck C.A., H.A.Maeda. 2018. Tyrosine biosynthesis, metabolism, and catabolism in plants. *Phytochemistry* 149: 82-102.

Teixidor-Toneu, I., F.M. Jordan, J.A. Hawkins. 2018. Comparative phylogenetic methods and the cultural evolution of medicinal plant use. *Nature Plants* 4: 754-761.

Ververidis F., E. Trantas, C. Douglas, G. Vollmer, G. Kretzschmar, N. Panopoulos. 2007. Biotechnology of flavonoids and other phenylpropanoid-derived natural products. Part 1: Chemical diversity, impacts on plant biology and human health. *Biotechnology Journal* 1: 1214-1234.

Chapter 1: Identification of herbs used in Hmong post-partum chicken soup via DNA barcoding

Alex H. Crum[a], Zongxee Lee[b], Mayyia Lee[b], Natalie Hoidal[c], Sophie Naylor[a,1], Adiel J. Andino-Acevedo[a], Annika Smeenk[a], Zachary J. Radford[a,3], Lindsey Miller[d], Ya Yang[a]

[a]Department of Plant and Microbial Biology, University of Minnesota - Twin Cities, 1445 Gortner Avenue, St. Paul, MN 55108, USA

[b]3850 White Bear Ave, White Bear Lake, MN 55110

[c]Department of Horticultural Science, University of Minnesota - Twin Cities, 1970 Folwell Avenue, St. Paul, MN 55108, USA

[d]Minnesota Landscape Arboretum, University of Minnesota Extension, 3675 Arboretum Boulevard, Chaska, MN 55318

Corresponding author: Alex H. Crum (a-crum@live.com)

[1]Present Address: Colgate University, 13 Oak Drive, Hamilton, NY, 13346

[2]Present Address: Department of Biological Sciences, University of South Carolina, 700 Sumter St #401, Columbia, SC 29208, USA

Abstract

Ethnopharmacological relevance

The Hmong people of Southeast Asia have a well-developed system of herbal medicine, which refugees brought with them around the world after the Secret War in Laos. Of particular importance is a postpartum chicken soup, comprising a customizable ingredient list of medicinal herbs boiled with chicken. These herbs are now grown and sold in places where Hmong people have settled, including Minnesota, USA. Hospitals in the area have also begun offering a version of the soup to patients. However, the herbs are difficult to identify, particularly in the non-native climate of Minnesota. To preserve the knowledge of these herbs, enable culturally competent healthcare, and further study herbal pharmacology, DNA-based identification of these herbs is needed.

Aim of the study

Here, we use sequences from the nuclear region ITS and chloroplast gene *rbcL* to identify species traditionally used in the chicken soup recipe of the Hmong community of Minnesota, with a focus on a family of traditional Hmong herbalists.

Materials and Methods

PCR of the selected regions was carried out for 39 chicken soup herbs grown by the Lee family or sold at a Hmong market. Sequences were used to query GenBank for similar sequences, and phylogenies of the genes were paired with morphology to identify species.

Results

Out of 38 herbs, we identify 30 to the species level. This is an improvement over past studies, which attempted to identify some of the same species identified here. Previous studies also had differences in specific ways the same herbs are used, as well as some names and species identifications. In the remaining species, identification was limited by taxonomic gaps in publicly available data, sequence quality, and taxonomic issues.

Conclusions

DNA barcoding enables identification of otherwise difficult to identify medicinal species. Additionally, differences compared to previous studies of Hmong medicinal herbs is likely due to a combination of misidentification, cultural drift, and diaspora-related changes.

1. Introduction

The Hmong people comprise an ethnic group that originated in China and spread throughout Southeast Asia. During the Vietnam War, a faction of Hmong living in Laos were recruited by the CIA to wage a secret war (1961–1975) combating the rise of communism in the country (Yang, 2003). After the Secret War's conclusion and rise of the communist government in Laos, thousands of Hmong were forced to flee the country to escape persecution, beginning a global diaspora (Yang, 2003). While Hmong refugees also settled in Europe, Australia, South America, and other Asian countries, the majority of these refugees settled in the U.S. Although initially Hmong refugees were dispersed throughout the country, over time much of the population has consolidated to live in Wisconsin, Minnesota, and California. As of 2020, it is estimated nearly 336,000 Hmong live in the U.S., with the second-largest population in Minnesota, numbering ~95,000 (Pfeifer, 2024).

In the diaspora, the Hmong brought much of their way of life with them, including a system of medicine and spirituality. Of particular importance in Hmong herbal medicine (tshuaj ntsaub), is the postpartum chicken soup (Jambunathan, 1995; Rice, 2000; Srithi et al. 2012). Traditionally, this soup, along with rice, is all that a person eats for the first 30 days after giving birth. Made with chicken and a suite of core herbs and more customizable additions, the soup is meant to strengthen and purify the body after the stress of giving birth (Jambunathan, 1995). Other restrictions, such as avoiding strenuous activity and confining themselves to the home, are also followed during this time to heal the birth giver's body and spirit (Jambunathan, 1995; Rice, 2009).

In the city of St. Paul, which hosts the largest population of Hmong in Minnesota, there has been recent interest in offering the chicken soup diet as part of culturally competent care in hospitals, with MHealth Fairview Birthplaces and Regions Hospital offering a version (Yuen, 2024; Vang, 2019). Sellers at Hmong markets in the city, such as Hmongtown Marketplace, also offer bundles of the core herbs for making the soup at home. However, there is also increasing concern that the tradition will be lost as younger generations of Hmong further integrate into American culture. Even outside the U.S., traditional Hmong medicinal knowledge is disappearing (Nguanchoo et al., 2022).

This concern is compounded by the fact that, while the Hmong names for the herbs (differing depending on whether the practitioner speaks the Green or White Hmong dialect) are generally well-known by those growing, selling, and prescribing them, the Latin names are often difficult to assign. Many practitioners of Hmong herbal medicine brought their herbs with them to the U.S., carefully stewarding them through long journeys and stays at refugee camps (Spring, 1989). Because Minnesota has a vastly different climate than highland Laos, the native habitat of

most of the herbs, many of the herbs rarely flower, and can have a different growth form than they would in their native habitat. Efforts to identify them are further stymied by limited English or Hmong-language keys. Previous efforts to identify the herbs have met with varied success. A previous catalog of the herbs being used by Hmong in Minnesota, was able to identify 20 out of the 52 herbs they investigated to species based on morphology, with the rest having tentative species identification, or only known to genus, family, or not at all (Spring, 1989). In Thailand, an investigation of 79 species used by Hmong for women's healthcare identified 71 to the species level, with the rest to genus or family (Srithi et al., 2012)**.**

Apart from the importance of preserving knowledge for future generations of Hmong, Latin names are essential for the continued offering and improvement of the postpartum chicken soup in hospitals. Currently, the soup offered by hospitals is limited to a few identified herbs, to avoid potential contraindications or safety issues.

Hospitals offering the chicken soup source their materials from local Hmong farmers, including the Lee family. Mayyia Lee, the matriarch of the family, was trained in herbal medicine in Laos by her mother, who served as a community healer. Mayyia brought her herbs with her when she came to the U.S. in 1980. The familial tradition is carried on by her children.

In an effort to preserve their family's knowledge, Mayyia's daughter Zongxee Lee initiated this work. Here, we use DNA barcoding methods to identify the herbs grown, used, and sold by the Lee family for postpartum chicken soup. Of the 39 herbs that can be used in the Lees' postpartum chicken soup recipe, 30 were identified to species, 5 to genus, and 4 to family. This is an expansive list that encompasses and surpasses the diversity of herbs sold for chicken soup at Hmongtown Marketplace when we visited, which is a more standardized "one-size-fits-all" recipe. The results showcase the power of molecular methods in identifying ethnobotanically

important plants, as well as the continued importance of publicly available molecular barcoding data.

2. Materials and Methods

2.1 Plant material collection

Plants were collected between July 2021 and June 2023 from 3 locations: May Lee's farm plot at Big River Farms (Marine on St. Croix, MN), Zongxee Lee's house (Roberts, WI), and Hmongtown Marketplace (St. Paul, MN). Plants collected at Big River and Zongxee Lee's house were grown by the Lee family, who confirmed their Hmong names. Plants collected from Hmongtown Marketplace were purchased from a vendor as a chicken soup herb bundle, and Hmong names were later confirmed by May and Zongxee Lee. For each plant, leaf tissue was placed in silica gel to be dried for DNA extraction, and a voucher specimen was taken. Photographs of the herbs were also taken to document morphology throughout the 2022 and 2023 growing seasons. Additionally, May and Zongxee Lee described the uses of herbs.

2.3 DNA extraction, PCR, and gel electrophoresis

DNA extraction from the silica-dried leaves was performed using the QIAGEN DNEasy Plant Mini Kit following the manufacturer's instructions. PCR amplification was first attempted with ITS, then *rbcL* if ITS amplification failed or more information was needed for identification. Primer sequences for ITS and *rbcL* amplification are given in Table 1.1. PCR amplification occurred in a 15 µL mix, with 0.1 µL Takara ExTaq, 0.5 µL of each primer, and 2 µL DNA template**.** DNA was diluted between 1:50 and 1:1000 to reduce enzyme inhibition by secondary metabolites. To troubleshoot the standard PCR protocols for resistant samples, 1µL

bovine serum albumin was added to PCR reactions, or the strength of the DNA dilution was altered. Success of PCR was checked with gel electrophoresis on a 1.5% agarose gel.

Table 1.1- Primer sequences used for PCR of medicinal herbs

Region	**Forward Primer**	**Reverse Primer**	**Source**
ITS	(ITS-I) 5'-GTC CAC TGA ACC TTA TCA TTT AG-3'	(ITS4) 5'-TCC TCC GCT TAT TGA TAT GC-3'	Urbatsch et al., 200 White et al., 1990
rbcL	5'-ATG TCA CCA CAA ACA GAA ACT AAAGC-3'	5'-CTT TTA GTA AAA GAT TGG GCC GAG-3'	Gilmore & Hill, 1997 Manhart, 1994

2.4 Sequence analysis, phylogenetic reconstruction, and taxonomic reconciliation

Successful PCR products were purified using ExoSAP-IT PCR Product Cleanup Reagent (Applied Biosystems) and forward and reverse strands were sent for Sanger sequencing by GENEWIZ from Azenta Life Sciences. The returned sequences were assembled in Geneious Prime v2023.1.1 (www.geneious.com). Within Geneious, candidate species identifications of each herb were performed via a BLASTn against the National Center for Biotechnology Information (NCBI)'s Nucleotide Collection (nr/nt) database. A subset of the top BLAST hits for each herb sequence were then aligned with the herb sequence via the MUSCLE plugin on Geneious (Edgar, 2004). Other databases, such as the Barcode of Life Data System v4 (BOLD, https://www.boldsystems.org/), were also queried, but the data retrieved was ultimately excluded, as much of it was either a duplication of a GenBank entry or did not align well with other sequences (Ratnasingham & Herbert 2007). Maximum-likelihood phylogenetic analysis plus 100 bootstrap replicates were carried out on each gene alignment via the RAxML v8.2.11 plugin in Geneious (Stamatakis, 2014). The sister species, phylogenetic resolution, and branch lengths of

the tree, as well as overall quality and length of sequences in the alignment, were all considered in determining the putative identity of the Hmong medicinal herb. As species names for sequences on GenBank may not always reflect the most current, accurate names and different submissions may have different names for the same species, putative names were checked using World Flora Online (www.worldfloraonline.org) and Plants of the World Online (www.powo.science.kew.org). Additionally, the native range of the putative species and/or possibility of introduction to SE Asia were checked in Plants of the World Online.

2.5 Morphological confirmation

Once a putative identification to species was made, the voucher specimen and photos of May and Zongxee Lee's herbs were compared to research grade images on iNaturalist, digitized herbarium specimens on the JSTOR Global Plants Database (plants.jstor.org), and descriptions from the Flora of China (Flora of China, 2015). Additionally, as May and Zongxee Lee have the greatest familiarity with the herbs, they verified the species determination by comparing images of the putative species with their knowledge of the herbs.

3. Results

Of the 39 herbs involved in postpartum chicken soup, we identified 30 to species, 5 to genus, and 4 to family (Table 1.2). The identification of some herbs, such as zab zi, were complicated by a lack of sequence data from closely related species. Others, notably tshab xyoob and moj tsuas in the Apiaceae family, were unable to be confidently identified due to apparent taxonomic issues in the family, with identical sequences falling under several accepted names or synonyms for different names and the tree being largely unresolved (Fig 1.1).

The herbs tauj dub (*Cymbopogon citratus*), kub muas lwj (*Kalanchoe laciniata*), and hmab ntsha ntsuab/liab *(Basella alba*) were identified based on scent/taste, leaf shape, and growth habit, being distinctive even without floral characteristics available. Similarly, although PCR was unsuccessful after multiple attempts at both loci, sam mos kab has been identified down to the family Crassulaceae, based on the succulent stems and simple stipule-less leaves. Evidence that supports each herb's identification is summarized in Table 1.2**.** Comparisons to previous studies of the Hmong pharmacopeia are summarized in Table 1.3.

Table 1.2- Hmong postpartum chicken soup herbs used by the Lee family, with uses and support for identifications. BS= bootstrap support

Hmong name used by Lee Family (Collection #)	**Use in Soup**	**Species Identity**	**Regions sequenced**	**BLAST Pairwise Identity Support for ID**	**Tree Support for ID**	**Morphology Evidence for ID**
Hmab Ntsha Liab (No voucher)	Breast milk production, digestion, headaches	*Basella alba* L. (Basellaceae)	Identified without PCR	NA	NA	Vining habit and glossy succulent leaves
Hmab Ntsha Ntsuab (No voucher)	Upset stomach, cholesterol, digestive health	*Basella alba* L. (Basellaceae)	Identified without PCR	NA	NA	Vining habit and glossy succulent leaves
Kab Raus Liab (Crum 38/Crum 51)	Immune system health, strength, endurance, respiratory and urinary infections, diuretic, laxative	*Houttuynia cordata* Thunb. (Saururaceae)	ITS	99.1%	Only BLAST matches returned for tree building were accessions of *H. cordata*	Cordate leaves with red margins and inflorescence a spike white petal-like bracts

Ko Taw Os Dawb (Crum 30)	Indigestion, improving blood circulation after heavy blood loss during labor	*Angelica decursiva* Franch. & Sav. (Apiaceae)	ITS, rbcl	100%/100%	United in a clade with *A. decursiva* in ITS tree. rbcL tree is largely unresolved	Purplish stem, with sheaths around petioles and pinnately compound leaves
Ko Taw Os Liab (Crum 10)	Regaining strength, endurance, and appetite, expelling lochia and improve uterine strength	*Artemisia lactiflora* Wall. Ex DC. (Asteraceae)	ITS	99.4% *Artemisia lactiflora* and *A. tangutica*	Forms a polytomy with *A. lactiflora* and a clade of *A. comaiensis* and *A. moorcroftiana*	Purple stems with pinnately compound leaves. Leaflets irregularly serrated
Kob Lij Xeeb (Crum 27)	Ulcers, diarrhea, appetite, and breast milk production	*Talinum paniculatum* (Jacq.) Gaertn. (Talinaceae)	rbcL	99.9%	United in a *T. paniculatum* clade with 100% support	Fleshy leaves, reddish stem, tiny pink flowers with yellow or red fruits that turn brown at maturity
Kua Txob Ntsuab (Crum 13/Crum 23)	Typically only used in chicken soup if the person is experiencing blood clots	*Dicliptera chinensis* (L.) Juss. (Acanthaceae)	ITS	93.7% (short sequence)	Forms a clade with *D. napiere,* sister to *D. chinensis*	Bears little resemblance to *D. napiere*. Instead ovate, glabrous leaves suggest *D. chinensis*
Kua Txob Liab (Crum 14)	Same as kua txob ntsuab	Same species or closely related to kua txob ntsuab (*Dicliptera sp.*)	Failed PCR	NA	NA	Looks identical to kua txob ntsuab, but tea made from it is red, rather than green like kua txob ntsuab tea

Kuab Nplai Dib (Crum 11)	Not typical in soup, but can treat nausea and morning sickness prepartum	*Sedum sarmentosum* Bunge (Crassulaceae)	ITS	100%	Nested in a clade of other *S. sarmentosum* sequences	Acute leaf apices distinguish *S. sarmentosum* from *S. emarginatum*
Kuab Nplai Taub (Crum 12)	Not typical in soup, but can treat nausea and morning sickness prepartum	*Sedum emarginatum* Migo (Crassulaceae)	ITS	98.3%		Rounded leaf apices distinguish *S. emarginatum* from *S. sarmentosum*
Kub Muas Lwj (Crum 48)	Stomach pain, heart burn, inflammation	*Kalanchoe laciniata* (L.) DC. (Crassulaceae)	ITS	No *K. laciniata* ITS sequences on GenBank	Nested in *Kalanchoe*	Succulent, extremely dissected leaves distinguish from the rest of *Kalanchoe*
Moj Tsuas (Crum 15)	Indigestion, heartburn, irregular periods, kidney health	Apiaceae sp.	ITS		Forms a polytomy to other identical sequences under different names	
Ncas Liab (Crum 40)	Not a common chicken soup herb, and generally considered risky, abortifacient, labor inducer, passing the placenta when it has failed to detach from the uterine wall	*Achyranthes bidentata* Blume (Amaranthaceae)	ITS	100%	Nested in a clade with other *Achyranthes bidentata* sequences (82 BS)	Red-tinged stems and opposite, elliptic red leaves

Nkaj Liab (Crum 39)	Expels lochia, wound healing	*Iresine diffusa* Humb. & Bonpl. ex Willd. (Amaranthaceae)	ITS	99.2%	Highly supported clade of *I. diffusa* sequences (97 BS)	Red, heart shaped leaves with visibly paler pinnate venation
Nplooj Thuaj Kau (Crum 16)	Nausea or lack of appetite, not a common chicken soup herb	*Kalanchoe pinnata* (Lam.) Pers. (Crassulaceae)	ITS	97%	Forms a clade with other *K. pinnata* sequences	Succulent with red stems and toothed leaves with red edges
Ntiv Dub (Crum 50)	Sometimes confused for Ntiv and sold in soup bundles, but is not itself a chicken soup herb as it can cause stomach upset.	*Ruellia simplex* C.Wright (Acanthaceae)	ITS	98.9%	Forms a polytomy at base of *Ruellia*	Purple flowers and narrow linear leaves distinguish from other similar *Ruellia sequences*
Ntiv (Crum 17)	General tonic, milk production, appetite promotion, nausea	*Eupatorium fortunei* Turcz. (Asteraceae)	ITS	97%	Forms a clade with other *E. fortunei* sequences (100 BS)	Similar in appearance to Minnesota native *Eutrochium purpureum*, but has two lobed leaves at each node, instead of 3-5.
Pawj Ia (Crum 36/Crum 45)	Flavoring, heartburn, constipation	*Acorus calamus* L. (Acoraceae)	ITS	100%	Nested in *A. calamus* clade	Leaves reddish at base
Pawj Qaib (C54)	Flavoring, expelling lochia and clots	*Acorus gramineus* Aiton	ITS	100%	Nested in *A. gramineus* clade	Leaves green all the way through and generally more narrow than *A. calamus*

Qhib Zeb (Crum 46)	Not often a chicken soup herb. Pregnancy related fatigue and morning sickness	*Sedum emarginatum* Migo (Crassulaceae)	ITS	98.5%	Forms a highly supported clade with other *S. emarginatum* sequences (100 BS)	Creeping habit with rounded leaf apices
Qhuas Rau Qaib (Crum 47)	Flavoring, milk production, aches and pains	*Hedychium coronarium* J.Koenig (Zingiberaceae)	ITS	100% with several *Hedychium* species	Forms a polytomy with several *Hedychium* species	White flowers with white filament distinguish from other *Hedychium* species
Quab Quav Yeev (Crum 9)	Indigestion or diarrhea	*Persicaria chinensis* (L.) H. Gross (Polygonaceae)	ITS, rbcL	99.6%/88.1%	Sister to a clade of *P. chinensis* and *P. microcephalum*	Leaves have two auricles at base of winged petioles, consistent with Flora of Pakistan description
Raws (Crum 34/Crum 35)	Uncommon in chicken soup, but can be added to reduce swelling in extremities or treat constipation	*Commelina communis* L. (Commelinaceae)	rbcL	99.6%	Nested in a monophyletic clade (99 BS) of other *C. communis* species.	2 showy blue petals, and one reduced white one
Sab Txhim Hmab (Crum 32)	Stomach ulcers, healing scar tissue	*Anredera cordifolia* (Ten.) Steenis (Basellaceae)	rbcL	96.2%	In a highly supported (97 BS) monophyletic clade with other *A. cordifolia sequences*	Vining habit, succulent leaves, and racemes of white flowers

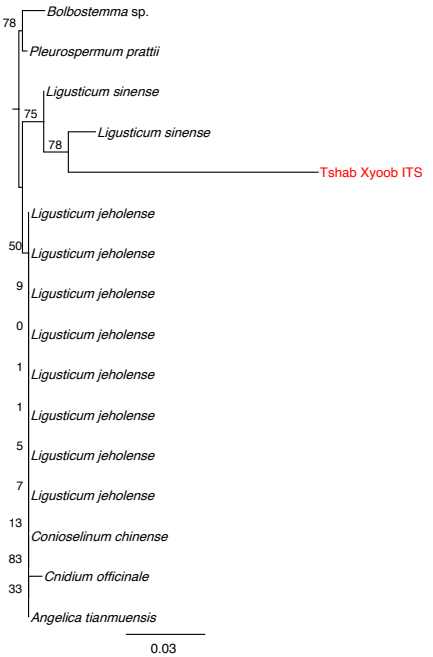
Sam Mos Kab (Crum 37)	Fatigue, ulcers, milk production, nausea	Crassulaceae sp.	Failed PCR	NA	NA	Vegetative growth similar to *Hylotelephium spectabile*
Siv Toj (Crum 43)	Rarely used due to smell. Uterine healing, stomach pain or ulcers	*Valeriana jatamansi* Jones (Caprifoliaceae)	ITS	99.6% with *V. officinalis, V. wallrothii,* and *V. californica*	No *V. jatamansi* ITS sequences to include in tree.	Simple leaves with crenulate margins and pubescent stems and leaves match the description in Flora of China
Suv Ntsim Iab (Crum 44)	Persistent digestive issues	*Artemisia verlotiorum* Lamotte (Asteraceae)	ITS	100%	Forms a polytomy with *A. verlotiorum , A. argyi,* and *A. igniaria*	Leaves more dissected than *A. argyi* and *A. igniaria*
Suv Ntsim Nyeg (Crum 31)	Interchangeable with Suv Ntsim	*Artemisia annua* L. (Asteraceae)	ITS	99.9%	Forms a well-supported (98 BS) clade with other *A. annua* species	Feathery, dissected leaves distinctly different than the other two suv ntsims.
Suv Ntsim (Crum 3)	Flavoring and detoxification	*Chrysanthemum indicum* L. (Asteraceae)	ITS, rbcL	99.8%/99.8%	Nested within clade of *C. indicum* and *C. lavandulifolium* sequences	Rhizomatous growth habit differentiates it from *C. lavandulifolium* (Flora of China)
Zab Zi (Crum 26)	Blood flow	Iridaceae sp.	ITS	88.2 % with *Tigridia mexicana* and *Cardiostigma longispatha*	Tree contains several polytomies	Leaves sprout from a red bulb. White flowers with free, symmetrical petals and yellow anthers

Tseej Ntug (Crum 49)	Aids digestion, constipation, blood circulation	*Silene viscidula* Franch. (Caryophyllaceae)	ITS	99.6%	Forms a clade (100 BS) with some *S. viscidula* sequences, but *S. viscidula* is polyphyletic	Elliptic leaves with very prominent midvein and an attenuate base matches Flora of China description
Tshab Xyoob (Crum 19/Crum 25)	Warming the body and circulation, uterine health	Apiaceae sp.	ITS		Forms a polytomy to other identical sequences under different names	
Tshais Qav (Crum 33)	Fevers, flavoring	*Aster* sp. (Closely related to *A. indicus*)	ITS, rbcL	100% *Aster indicus, Kalimeris indica, Miyamayomena piccolii*/99.7% with several *Aster* species	Forms a polytomy with other *Aster* species, including indicus, in both trees	Light lavender rays, yellow disk flowers, oblanceolate leaves
Tshuaj Roj Liab (Crum 4)	Endurance. Blood circulation, post-labor recovery	*Gynura* sp. (Asteraceae)	rbcL	99.8% to *G. japonica* and *Senecio* sp.	High support within the *Gynura* clade, but unresolved within	Similar in appearance to *G. bicolor,* with purple abaxial leaf surface, but covered in dense hair
Tshuaj Roj Tsuab Soob (Crum 6)	Swelling, bringing the fat to the top of the soup	*Gynura procumbens* (Lour.) Merr. (Asteraceae)	ITS	99.8%	Forms a clade with *G. procumbens* sister to the rest of *Gynura;* 84 bootstrap suppor	Distinguished from the other *Gynura herbs* by sparser hair and both leaf surfaces green

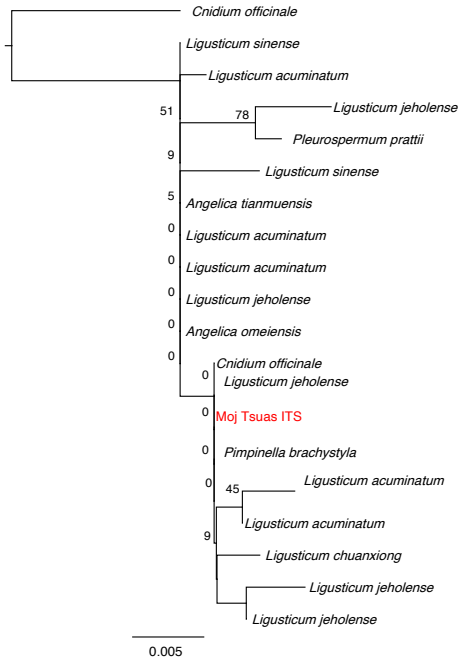
Tshuaj Roj Tsuab (Crum 5)	Inflammation bringing the fat to the top of the soup	*Gynura* sp. (Asteraceae)	ITS	100% to *Senecio* sp. and 99.9% to *G. japonica*	High support within the *Gynura* clade, but unresolved within	Similar in appearance to *G. bicolor,* but covered in dense hair and leaves are green to only slightly purple
Tuaj Dub (No voucher)	Flavoring, digestion, and bladder function	*Cymbopogon citratus* (DC.) Stapf (Poaceae)	NA	NA	NA	Identified by aroma and flavor
Zaj Xaws (Crum 7)	Swelling, bringing the fat to the top of the soup	*Gynura sp.* (Asteraceae)	ITS, rbcL	98.6% *G. bicolor*/99.3% *G japonica*	Nested in polytomies of multiple *Gynura* species	Likely *Gynura bicolor* based on glabrous entire leaves with purple abaxial surface
Zej Nthsua Ntaug (Crum 8)	Endurance	*Elsholtzia penduliflora* W.W. Sm. (Lamiaceae)	ITS	100%	Forms a well-supported (100 BS) clade with other *E. penduliflora* sequences	Square, ridged stems with purple running down the center and frequent purple bumps. Opposite leaves with serrated purple margins

Fig. 1.1- a-b**)** Maximum likelihood phylogenies of ITS of two Apiaceae herbs, tshab xyoob and moj tsuas (highlighted in red) to illustrate the polytomies caused by identical and near-identical sequences under different names. c) Maximum likelihood phylogeny of *rbcL* of kob lij xeeb to illustrate a positive, well-supported identification where the herb ITS sequence forms a monophyletic group with other accessions of the species. Numbers at nodes represent bootstrap support. Scale bars represent branch length. Trees are rooted by the most distantly related taxa in the tree. d) Images of the herbs clockwise starting from top left: Tshab xyoob, moj tsuas, kob lij xeeb. Photo credits Lindsey Miller, Anna Andreasen, Alex Crum

a)


Bolbostemma sp.
Pleurospermum prattii
Ligusticum sinense
Ligusticum sinense
Tshab Xyoob ITS
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Ligusticum jeholense
Conioselinum chinense
Cnidium officinale
Angelica tianmuensis
78
75
78
50
9
0
1
1
5
7
13
83
33
0.03

b)


Cnidium officinale
Ligusticum sinense
Ligusticum acuminatum
Ligusticum jeholense
Pleurospermum prattii
Ligusticum sinense
Angelica tianmuensis
Ligusticum acuminatum
Ligusticum acuminatum
Ligusticum jeholense
Angelica omeiensis
Cnidium officinale
Ligusticum jeholense
Moj Tsuas ITS
Pimpinella brachystyla
Ligusticum acuminatum
Ligusticum acuminatum
Ligusticum chuanxiong
Ligusticum jeholense
Ligusticum jeholense
51
78
9
5
0
0
0
0
0
0
0
45
9
0.005

c)

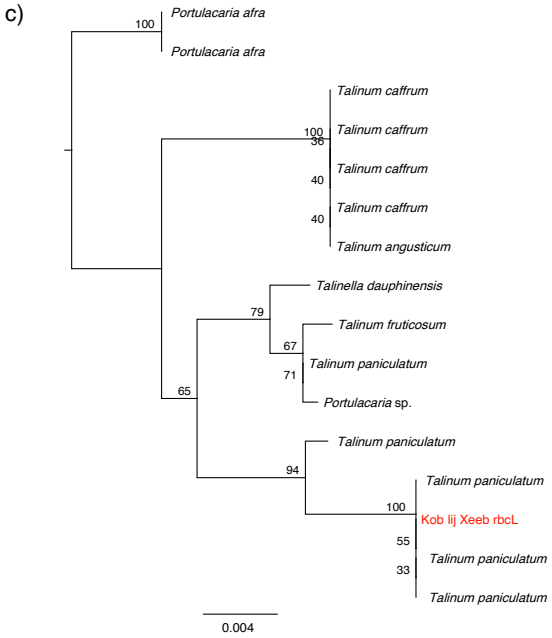

Portulacaria afra
Portulacaria afra
Talinum caffrum
Talinum caffrum
Talinum caffrum
Talinum caffrum
Talinum angusticum
Talinella dauphinensis
Talinum fruticosum
Talinum paniculatum
Portulacaria sp.
Talinum paniculatum
Talinum paniculatum
Kob lij Xeeb rbcL
Talinum paniculatum
Talinum paniculatum
100
100
36
40
40
79
67
71
65
94
100
55
33
0.004

Table 1.3- Comparison of Hmong herbs overlapping with herbs cataloged in previous studies in Minnesota (Spring, 1989), California (Corlett et al., 2003, 2009), and Thai (Srithi et al., 2012) communities.

Hmong name as given by Lee Family	**Species identity in current study**	**Comparison with existing literature**
Hmab ntsha ntsuab/liab	*Basella alba*	Called Hmab ntsha liab in Srithi et al. 2012, as Hmab Ntsha in Corlett et al. 2003 and 2009, and maab ntshaa (Green dialect only) in Spring 1989.
Kab raus liab	*Houttuynia cordata*	Called Kab raus in Spring 1989 and Srithi et al. 2012. And Kab tsaus in Corlett et al. 2003, 2009
Ko taw os dawb	*Angelica decursiva*	In Corlett et al. 2009 & 2003, a herb called Ko taw os is identified as aff. *Angelica*
Ko taw os liab	*Artemisia lactiflora*	*Artemisia lactiflora* is called Tab kib liab luj in Srithi et al, but still has the same postpartum use
Kob lij xeeb	*Talinum paniculatum*	Called nkob lij xeeb in Spring 1989.
Kua txob ntsuab/liab	*Dicliptera chinensis*	A *Dicliptera* sp. is named Tshuaj kua txob in Srithi et al. 2012, and is used similarly.
Kuab nplai dib	*Sedum sarmentosum*	Called Nplai zeb and identified as *Sedum* cf. *sarmentosum* in Srithi et al. 2012. In Corlett et al. 2009 & 2003, called kuab nplais dib and identified as *Sedum* aff. *sarmentosum*
Kub muas lwj	*Kalanchoe laciniata*	*Kalanchoe laciniata* is called Tshuaj ntiv in Srithi et al. 2012
Nkaj liab	*Iresine diffusa*	Identified as *Iresine herbstii* Srithi et al. 2012 and Spring 1989
Nplooj thuaj kau	*Kalanchoe pinnata*	Called Nplooj tuaj kaus (Green dialect only) in Srithi et al 2012. Called nplooj tuaj daus and identified as *Kalanchoe* cf. pinnata in Spring 1989
Ntiv	*Eupatorium fortunei*	Identified as *E. lindleyana* n Corlett at el. 2009 & 2003(An invalid name probably meant to indicate *E.*

		lindleyanum). In Spring 1989, Ntiv is identified as *Valeriana* cf. *officinalis*
Pawj ia	*Acorus calamus*	Also identified as *A. calamus* in Srithi et al. 2012
Pawj qaib	*Acorus gramineus*	Also identified as *A. gramineus* in Srithi et al. 2012, Corlett et al. 2009, Corlett et al. 2003, and Spring 1989
Quab quav yeev	*Persicaria chinensis*	An herb called Guab quav yeeb is identified as *Polygonum runcinatum* in Spring 1989.
Raws	*Commelina communis*	An herb called raws ntsuab is identified as *Commelina* cf. *communis* in Spring et al. 1989
Sab txhim hmab	*Anredera cordifolia*	Also identified as *A. cordifolia* in Srithi et al. 2012, and Spring 1989
Sam mos kab	Crassulaceae sp.	Called Sam muaj kab (Green dialect only) in Srithi et al. 2012 and Sam Moj dab in Spring 1989. Identified in each as Crassulaceae sp. and *Sedum telephium,* respectively.
Suv ntsim	*Chrysanthemum indicum*	Suv Ntsim identified as *Artemisia vulgaris* in Srithi et al. 2012
Tshab xyoob	Apiaceae sp.	Mentioned as a unidentifiable Apiaceae species in Srithi et al. 2012
Tshuaj roj liab, Tshuaj roj ntsuab, Tshuaj roj ntsuab soob	*Gynura* spp.	An herb called Tshuaj rog is identified as *Gynura bicolor* in Srithi et al., 2012.

4. Discussion

4.1 Customizability of the chicken soup recipe

Two core herbs are always included in postpartum chicken soup: ko taw os liab (*Artemisia lactiflora)*, and ntiv (*Eupatorium fortunei*). In addition, at least one of the four *Gynura* herbs, tshuaj roj liab, tshuaj roj ntsuab, tshuaj roj ntsuab soob (*Gynura procumbens*), and

zaj xaws, are typically added to bring the chicken fat to the broth's surface to be skimmed off while cooking. Ntiv dub (*Ruellia simplex*), while not a typical soup herb, is included here because it is commonly confused with the more typical herb ntiv and can have deleterious effects such as stomach pain. The remaining herbs can be added to the soup in different numbers and combinations depending on the recipient's postpartum symptoms or to vary the flavor. For example, raws (*Commelina communis*), while not often included, can be added to the soup if the recipient has edema in their limb extremities. Suv ntsim (*Chrysanthemum indicum*), zej ntshua ntuag (*Elsholtzia penduliflora)*, kuab nplai taub (*Sedum emarginatum*), kuab nplai dib (*Sedum sarmentosum*), tuaj dub (*Cymbopogon citratus*), and pawj qaib (*Acorus gramineus*) are among the herbs most commonly added.

4.2 Barcoding as a tool for medicinal herb identification

To our knowledge, this is the first study to incorporate DNA to identify plants in the Hmong pharmacopeia. We were able to use the ITS and *rbcL* regions to identify the majority of the herbs to species. Even when molecular data were unable to resolve an herb to species, in many cases, (i.e. kua txob ntsuab and suv ntsim iab), they often provided enough information to narrow down choices and then make an ID based on morphology.

However, several challenges became apparent in the process of DNA barcoding. Some herbs, such as kua txob ntsuab, consistently yielded short or low-quality ITS sequences despite multiple attempts. In the case of kua txob ntsuab, however, despite amplification of *rbcL* repeatedly failing, the short ITS sequence was still enough to identify a genus, which vegetative morphology alone was unable to (Table 1.2, Fig. S1.1).

Other species identifications were complicated by misidentifications in the database and/or taxonomic issues. This is most apparent in the unidentified Apiaceae herbs, Tshab Xyoob and Moj Tsuas. Both these herbs had several top BLAST matches that were identical to each other, despite being attributed to different species and even different genera within Apiaceae. As a result, the trees were poorly resolved, with the sampled herb sequence nested in a polytomy with different species names (Fig. 1.1). This is either a widespread case of misidentifications in the database entries, or the same species existing under multiple accepted names. The taxonomy of *Ligusticum* and its allies is known to be problematic, with a significant degree of polyphyly and homoplasy (Zhou et al,. 2020). Further complicating classification in this group is the lack of informative sites in regions typically used for barcoding and phylogenetics, including ITS (Wei et al., 2022; Zhou et al., 2020).

In still other cases, such as zab zi and the *Gynura* herbs, identification was hindered by a lack of publicly available ITS or *rbcL* data in the family or genus, resulting in few close BLAST matches and unresolved phylogenies (Fig S1.2-S1.6). While the trend in plant biodiversity research has moved towards increasingly powerful whole-genome or transcriptome methods, Sanger sequencing remains an important tool for barcoding studies like this one, where identification based on a whole genome is financially and computationally infeasible. Increasing the taxonomic coverage of barcoding loci in publicly available databases is still crucial.

4.3 Comparison to previous studies in Hmong communities

Our results are partially congruent with previous studies of Hmong medicinal herbs. 20 herbs overlapped between previous work, although at times the Hmong names of an identified species or the identified species ascribed to a Hmong name differed (Corlett et al., 2003; Corlett et al, 2009; Srithi et al., 2012; Spring, 1989). These other studies were carried out in Minnesota

(Spring, 1989), California (Corlett et al., 2003; 2009), and Thailand (Srithi et al., 2012). Only the study in Thailand specifically focused on herbs used for women's healthcare (Srithi et al., 2012). The other studies, conducted in the United States by Western scientists, were more general, cataloging the medicinal herbs used and grown by Hmong communities in the U.S. (Corlett et al., 2003; Corlett et al, 2009; Spring, 1989). The American studies' informants were all women, and the Thai study included 5 women and 1 man. A comparison among studies is given in Table 1.3.

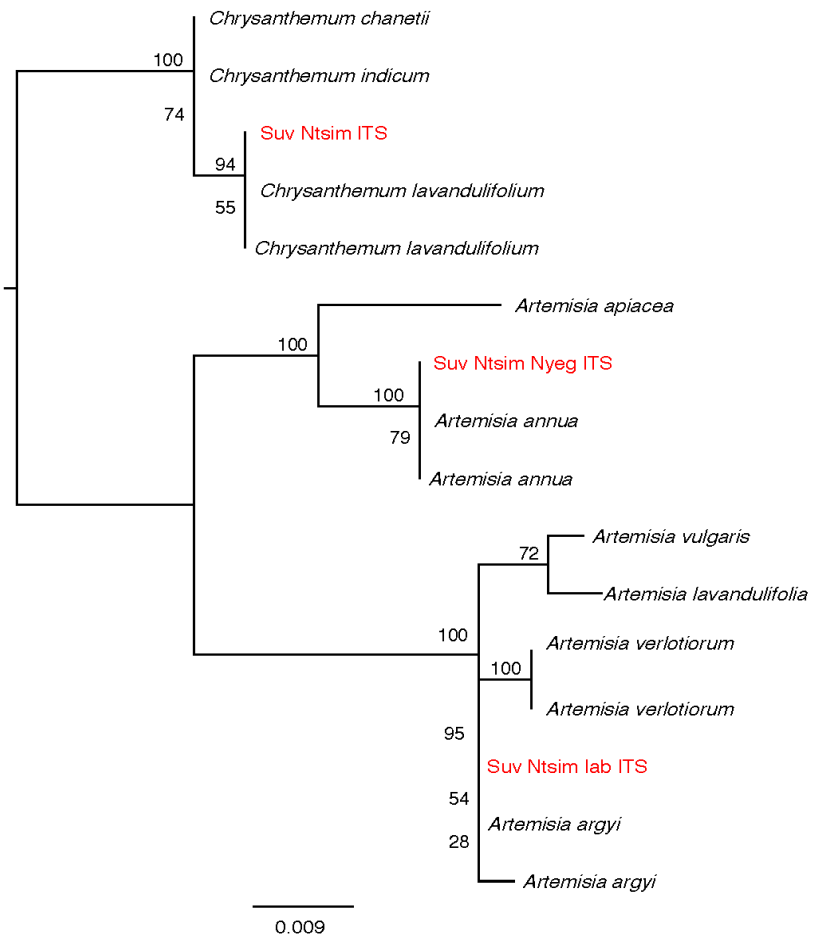
The Lee family speaks the White Hmong dialect, and in general, the names identified in this study typically matched the White Hmong names from previous studies. Several spelling differences in Hmong names occurred between studies, however, this may be due to difficulty in interpreting the tonal language into the phonetic signifiers of the Hmong Romanized Popular Alphabet (the most commonly used way of representing the Hmong language in text in the West; Table 1.3). Some Hmong names in previous studies also are sometimes shorter or less specific, often dropping a final word describing color. For example, what the Lee family calls kab raus liab (*Houttuynia cordata*) is referred to simply as kab ruas or kab tsuas in other works, dropping the Hmong word for "red" (Corlett, et al. 2003, 2009; Spring, 1989; Srithi et al. 2012). In addition, the herbs hmab ntsha and suv ntsim are further specified and divided by the Lee family. Red cultivars of *Basella alba* are called hmab ntsha liab, and green cultivars are called hmab ntsha ntsuab, and have slightly different uses.

Similarly, in Srithi et al. (2012), only a single *Gynura* herb is identified as *G. bicolor,* and is called tshuaj rog. However, here, we distinguish between tshuaj roj liab, tshuaj roj ntsuab, and tshuaj roj ntsuab soob, all of which are distinguishable from each other by leaf shape and hair. There is relatively little publicly available *Gynura* sequence data, and phylogenetic trees for ITS

and *rbcL* were largely unresolved, so it is unclear whether the more specific names represent different cultivars, as in *Basella*, or different species in the genus.

Suv ntsim is an even more complicated case. May believed that both suv ntsim and suv ntsim nyeg were the same species, with the "nyeg" ending indicating a plant grown in cultivation. Suv ntsim iab, on the other hand, was believed by May to be a different herb that only grew in the wild. However, our suv ntsim and suv ntsim nyeg sequences suggested that they were different species. In fact, suv ntsim is in an entirely different genus (*Chrysanthemum*) than suv ntsim nyeg and suv ntsim iab *(Artemisia;* Fig 1.2). Interestingly, suv ntsim in Srithi et al. (2012) is identified as *Artemisia vulgaris* rather than *Chrysanthemum indicum*.

Fig. 1.2- Phylogeny of the three suv ntsim herbs and a selection of their top BLAST matches, supporting the placement of suv ntsim in *Chrysanthemum* and the separate identities of suv ntsim and suv ntsim nyeg. Numbers on branches are the bootstrap support. Scale bar represents branch length.



Suv ntsim is not the only herb name with different species assigned to it. Ntiv, here identified as *Eupatorium fortunei,* is identified as the related *E. lindleyanum* in Corlett et al. (2003, 2009). However, in Spring (1989), ntiv is identified as *Valeriana* cf. *officinalis*. Given the dissimilarity in the appearance of *E. fortunei* and *V. officinalis* (i.e. simple vs. compound leaves), it is likely that the Minnesota Hmong surveyed in Spring (1989) call a different species ntiv than the one from the Lee family and sold by a different vendor at the Hmongtown Market. Additionally, while many of the postpartum herbs are repeatedly mentioned as such, specific medicinal use attributed to an herb varies substantially from study to study.

Although some of the differences may come down to misidentification based on morphology, another likely explanation is cultural variation. This work focuses on the herbs of one family, living in Minnesota since 1980. Spring (1989) targets an earlier, broader section of the same community in Minnesota, while Corlett et al. (2003, 2009) survey the California Hmong community. Srithi et al. (2012) surveys a Thai Hmong community.

4.4 Forces driving changing species use

Even before the Secret War drove a diaspora, the Hmong spread throughout Southeastern Asia from an origin in China. Post-diaspora, the Hmong are distributed globally. Changes in pharmacopeias may be due to species substitution, or introduction. In substitution, similar, closely related plants in the new location replace the original medicinal plant (Voeks, 2016). Substitution occurring in the Hmong community is supported by May's observation that, in the case of nkaj liab *(Iresine diffusa),* she has seen other sellers at Hmong markets selling a similar-looking herb that is not nkaj liab in place of it for chicken soup. This herb, which we also sequenced, was identified as *I. herbstii,* which is sometimes treated in synonymy or as a growth form or subspecies of *I. diffusa (*www.worldfloraonline.org)*,* In our tree, the two were closely related with ~93% sequence identity, suggesting that they are separate species (Fig. S1.7). *Iresine herbstii* is native to Peru and is widely cultivated as an ornamental, making it readily available in the U.S (powo.science.kew.org).

Another potential example of substitution is the occasional sale of ntiv dub instead of ntiv. Ntiv dub, despite not having the same reported effect as ntiv, looks somewhat similar when neither is flowering. Ntiv dub is *Ruellia simplex,* which is native to South America and a common ornamental plant in the U.S.

Some species were probably also introduced. For example, kob lib xeeb *(Talinum paniculatum*) is native to the Americas, but was probably introduced to the Hmong pharmaocopeia before the Secret War, given its presence in Asia and use in other Southeastern Asian cultures (Gu et al., 2020; Hong et al. 2015)

Differences in herbal use and identity between communities may also be explainable by cultural drift, where cultural differences arise in populations and eventually become fixed in a process analogous to genetic drift (Mesoudi, 2011). According to this theory, cultural traits such as medicinal plant use vary like alleles, with variation in a population that can become fixed over time. Different populations, or communities, can experience shifts in medicinal plant usage as differences arise and either die out or become fixed in that community. Although there is contact between California and Minnesota Hmong and, to a lesser extent, between Thai and American Hmong, even in the absence of cultural selective pressure, differences would accumulate over time between communities. This potentially accounts for differences between our study and the previous survey of Hmong living in Minnesota (Spring, 1989). Alternatively, differences in ease of cultivation based on climate or land usage in different regions may act as a selective pressure driving adaptation.

4.5 Community Impacts

The major goals of this work were 1) to preserve medicinal herb knowledge for future generations of Hmong living in the U.S. and 2) to provide species identity information to hospitals working towards culturally competent care that allows them to gauge the safety of the herbs. The species names of these herbs obtained through this study have allowed a review of their known pharmacology. In addition, the results of the DNA barcoding are being incorporated into a memoir and recipe book that is in the process of being published.

5. Conclusions

Here, we provide the first survey of Hmong post-partum herbs using DNA barcoding. DNA barcoding allows for greater certainty of a species identification than morphology alone, especially for herbs in a non-native climate where they rarely flower. Sequencing and publication of 1–2 loci associated with rigorous species identification and voucher information continue to be important for difficult identifications like the ones presented here. Differences in previous reports of medicinal herbs used by Hmong communities may be driven by a combination of diaspora-related changes, cultural drift, and misidentification. A wider survey of use within the Minnesota Hmong community and globally would begin to detangle these forces from each other.

6. Supplemental Materials

Figure S1.1 – Kua txob ntsuab ITS gene tree. For S1.1 and all subsequent suplemental figures, numbers at nodes represent bootstrap values

Figure S1.2 – Zab zi ITS gene tree.

Figure S1.3 – Tshauj roj liab rbcl gene tree

Figure S1.4 – Tshuaj roj ntsuab rbcL gene tree

Figure S1.5 – Tshuaj roj tsuab soob ITS gene tree

Figure S1.6 – a) ITS gene tree for zaj xaws b) rbcL zaj xaws gene tree

Figure S1.7 – ITS gene tree of nkaj liab purchased at Hmongtown Market and collected from Mayyia Lee.

7. References

Corlett, J.L., Dean, E.A., Grivetti, L.E. 2003. Hmong gardens: Botanical diversity in an urban setting. Economic Botany. 57, 365-379.

Corlett, J.L., Clegg, M.S., Keen, C.L., Grivetti, L.E. 2009. Mineral content of culinary and medicinal plants cultivated by Hmong refugees living in Sacramento, California. International Journal of Food Sciences and Nutrition. 53, 117-128. https://doi.org/10.1080/09637480220132139.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792-1797. https://doi.org/10.1093/nar/gkh340.

Editorial Committee of Flora of China. 2015. Chinese Academy of Sciences, Science Press, Beijing.

Gilmore, S., Hill, K.D. 1997. Relationships of the Wollemi Pine (*Wollemia nobilis*) and a molecular phylogeny of the Araucariaceae. Telopea 7, 275-291.

Gu, W., Hao, X., Wang, Z., Zhang, J., Huang, L., Pei, S. 2020. Ethnobotanical study on medicinal plants from the Dragon Boat Festival herbal markets of Qianxinan, southwestern Guizhou, China. Plant Diversity. 42, 427-433. https://doi.org/10.1016/j.pld.2020.12.010.

Hong, L., Gui, Z., Huang, K., Wei, S., Liu, B., Meng, S., Long, C. 2015. Ethnobotanical study on medicinal plants used by Maonan people in China. Journal of Ethnobiology and Ethnomedicine. 11, 32. https://doi.org/10.1186/s13002-015-0019-1.

iNaturalist. 2024. https://www.inaturalist.org/. (accessed 12 May 2024).

Jambunathan, J. 1995. Hmong cultural practices and beliefs: the postpartum period. Clinical Nursing Research. 4, 335-345. https://doi.org/10.1177/105477389500400309.

JSTOR Global Plants. 2024. https://plants.jstor.org/. (accessed 17 May 2024).

Manhart, J.R. 1994. Phylogenetic analysis of green plant rbcL sequences. Molecular Phylogenetics and Evolution. 3, 114-127.

Mesoudi, A. 2011. Cultural Evolution: How Darwinian theory can explain human culture and synthesize the social sciences. University of Chicago Press, Chicago.

Nguanchoo, V., Wangpakapattanawong, P., Balslev, H. Inta, A. 2022. Hmong medicinal plant knowledge transmission and retention in social modernity. Human Ecology 50, 419-433. https://doi.org/10.1007/s10745-022-00326-4.

Pfeifer, M. 2024. Hmong population trends in the 2020 U.S. census. Hmong Studies Journal. 26, 1-12.

Plants of the World Online. 2024. https://powo.science.kew.org/. (accessed 12 May 2024).

Ratnasingham, S. & Herbert, P.D.N. 2007. BOLD: The Barcode of Life Data System (www. barcodinglife.org). Molecular Ecology Notes. 7, 355-364. DOI:10.1111/j.1471-8286.2006.01678.x.

Rice, P.L. 2000. Nyo dua hli – 30 days confinement: traditions and changed childbearing beliefs and practices among Hmong women in Australia. Midwifery. 16, 22-34. DOI: 10.1054/midw.1999.0180.

Spring, M. 1989. Ethnopharmacologic analysis of medicinal plants used by Laotian Hmong refugees in Minnesota. Journal of Ethnopharmacology. 26, 65-91.

Srithi et al. 2012. Medicinal plants used in Hmong women's healthcare in northern Thailand. Journal of Ethnopharmacology. 139, 119-135.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30, 1312-1313. https://doi.org/10.1093/bioinformatics/btu033.

Urbatsch, L.E., Baldwin, B.G., Donoghue, M.J. 2000. Phylogeny of the coneflowers and relatives (Heliantheae:Asteraceae) based on nuclear rDNA internal transcribed spacer (ITS) sequences and chloroplast DNA restriction site data. Systematic Botany 25, 539-565.

Vang, G. Kare11 News. Regions Hospital to offer a taste of home to new Hmong moms. https://www.kare11.com/article/news/local/kare11-sunrise/regions-hospital-to-offer-a-taste-of-home-to-new-hmong-moms/89-852a13c7-0473-450d-a084-ede6c4a17e81. (accessed 17 May 2024).

Voeks, R.A. 2016. Diaspora Ethnobiology. In Albuquerque, U., Nóbrega Alves, R. (eds)

Introduction to Ethnobiology. Springer, Cham. https://doi.org/10.1007/978-3-319 28155-1 _7.

Wei, X.P., Zhang, X.Y., Dong Y.Q., Cheng, J.L., Bai Y.J., Liu, J.S., Qi, Y.D., Zhang, B.G., Liu, H.T. Molecular structure and phylogenetic analyses of the complete chloroplast genomes of three medicinal plants *Conioselinum vaginatum, Ligusticum sinense,* and *Ligusticum jeholense*. Front. Plant Sci. 13, 878263. DOI: 10.3389/fpls.2022.878263.

White, T.J., Bruns, T., Lee, S. Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In PCR Protocols: a guide to methods and applications.https://doi.org/10.1016/B978-0-12-372180-8.50042-1.

WFO. 2024. World Flora Online. http://www.worldfloraonline.org. (accessed 12 May 2024)

Yang, K., 2003. Hmong diaspora of the post-war period. Asian and Pacific Migration Journal. 12, 271-300. https://doi.org/10.1177/011719680301200302.

Yuen, L. Star Tribune. This herbal chicken soup has been nourishing Hmong moms for centuries. https://www.startribune.com/hmong-postpartum-soup-chicken-herb-moms-diane-moua-minnesota-laura-yuen/600365798. (accessed May 31, 2024).

Zhou, J., Gao, Y.Z., Wei, J., Liu Z.W., Downie, S.R. 2020. Molecular phylogenetic of *Ligusticum* (Apiaceae) based on nrDNA ITS sequences: rampant polyphyly, placement of the Chinese endemic species, and a much-reduced circumscription of the genus. Int. J. Plant Sci. 181, 306-323. DOI: 10.1086/706851

Chapter 2: Traditional medicinal use is linked with apparency, not specialized metabolite profiles in the order Caryophyllales

Alex H. Crum[1,4], Lisa Philander[2], Lucas Busta[3], Ya Yang[1]

[1]Department of Plant and Microbial Biology, University of Minnesota - Twin Cities, 1445 Gortner Avenue, St. Paul, MN 55108, USA

[2]Como Park Zoo and Conservatory, 1225 Estabrook Drive, St. Paul, MN 55103

[3]University of Minnesota - Duluth, 1038 University Drive, Duluth, MN 55812, USA

[4]Author for correspondence: Alex H. Crum a-crum@live.com

ABSTRACT:

Premise—Better understanding of the relationship between plant specialized metabolism and traditional medicine has the potential to aid in bioprospecting and untangling of cross-cultural use patterns. However, given the limited information available for metabolites in most plant species, understanding medicinal use-metabolite relationships can be difficult. The order Caryophyllales has a unique pattern of lineages of tyrosine- or phenylalanine-dominated specialized metabolism, represented by mutually exclusive anthocyanin and betalain pigments, making Caryophyllales a compelling system to explore the relationship between medicine and metabolites by using pigment as a proxy for dominant metabolism.

Methods—We compiled a list of medicinal species in select tyrosine- or phenylalanine-dominant families of Caryophyllales (Nepenthaceae, Polygonaceae, Simmondsiaceae, Microteaceae, Caryophyllaceae, Amaranthaceae, Limeaceae, Molluginaceae, Portulacaceae, Cactaceae, and Nyctaginaceae) by searching scientific literature until no new uses were recovered. We then tested for phylogenetic clustering of uses using a “hot nodes" approach. To test potential non-metabolite drivers of medicinal use, like how often humans encounter a species (apparency), we repeated the analysis using only North American species across the entire order and performed phylogenetic generalized least squares regression (PGLS) with occurrence data from the Global Biodiversity Information Facility (GBIF).

Key Results—We hypothesized families with tyrosine-enriched metabolism would show clustering of different types of medicinal use compared to phenylalanine-enriched metabolism. Instead, wide-ranging, apparent clades in Polygonaceae and Amaranthaceae are overrepresented across nearly all types of medicinal use.

Conclusions—Our results suggest that apparency is a better predictor of medicinal use than metabolism, although metabolism type may still be a contributing factor.

INTRODUCTION:

Nearly 80% of the population in low and middle income countries relies on medicinal plants as their primary form of healthcare (Gaoue et al., 2021; Hamilton, 2004) and globally, many will access pharmaceuticals developed from plant-based natural product extracts. Natural products in pharmacological studies are also known as specialized metabolites (SMs), reflecting the chemicals' function *in planta* rather than their service to people. SMs are important for plant survival, helping mediate biotic and abiotic stress responses and playing a role in plant communications (e.g. attracting pollinators). These compounds are extremely diverse, and it is estimated that the plant kingdom produces over one million unique SMs (Afendi et al., 2012). The specialized metabolite profile of a species tends to be influenced by phylogeny – while some SMs are relatively widespread, such as anthocyanins, others are more lineage-specific, such as the glucosinolates, whose only known occurrence outside Brassicales is in the family Putranjivaceae (Rodman et al., 1998). Overall, closely related plants tend to produce similar classes and abundances of SMs, being constrained by the SM biosynthetic pathways of their common ancestor (Holeski et al., 2021; Wink, 2003; Youssef et al., 2023).

Species used in traditional medicine are also clustered into certain phylogenetic lineages, with a species tending to be related to plants with similar use (Gaoue et al., 2021; Saslis-Lagoudakis et al., 2011, 2012). Because medicinal effect is determined by specialized metabolism and closely related plants tend to produce similar SMs, this phylogenetic pattern of traditional use likely reflects phylogenetic clustering of medicinal SMs (Pellicer et al., 2018). If

this hypothesis is correct, then traditional medicinal use in a clade should reflect the classes of SMs to be found in that clade. Though this correspondence can be clouded by instances of independent evolution, such as that observed for glucosinolates (Rodman et al., 1998; Huang et al., 2016; Yang et al., 2020), this approach has the potential to guide bioprospecting efforts, specifically as the gaps in reported knowledge in both ethnobotanical uses and metabolomics/natural products studies meet renewed interest in natural products research, in part to combat antibiotic resistant pathogens (Atanasov et al., 2021).

Although most of the challenges of associating SM and traditional medicinal use across a phylogeny cannot be dealt with in a single study, the order Caryophyllales is a compelling system to examine the relationship between SMs and traditional medicine due to its unique evolution of tyrosine- and phenylalanine-dominated specialized metabolisms. With approximately 12,500 species across 39 families, Caryophyllales includes many well-known species used in traditional medicine including: jojoba (*Simmondsia chinensis* (Link) C.K. Schneid*; Simmondsiaceae*), peyote (*Lophophora williamsii* (Lem. Ex. J.F. Cels) J.M. Coult.; Cactaceae), and Dock (*Rumex* spp.; Polygonaceae). Although suffering from the same lack of known metabolites as all other plant groups, the active constituents of some of the order's more charismatic plants have been identified. For example, mescaline from peyote is synthesized from tyrosine (Tyr) and is responsible for the plant's hallucinogenic effect.

Mescaline is also emblematic of a unique evolution of specialized metabolism in the order, which is also responsible for the Caryophyllales-specific betalain pigments. Mescaline, dopamine, and norepinephrine, together with the betalain pigments, which range from yellow to pink, are derived from the amino acid Tyr, and are part of a Tyr-derived SM diversification in the order (Chen et al., 2003; Lopez-Nieves et al., 2018). This diversification is the result of the

duplication of arogenate dehydrogenase (ADH), the gene responsible for converting Tyr from its precursor, arogenate, early in Caryophyllales (Lopez-Nieves et al., 2018). While the canonical version (ADH-β) is feedback-inhibited by excess tyrosine, the new copy (ADH-α) has relaxed that regulation. Species with ADH-α therefore produce an excess of tyrosine, which in Caryophyllales has been incorporated into specialized metabolism in a variety of ways, including isoquinoline alkaloids such as mescaline, betalains, and catecholamines (Lopez-Nieves et al., 2018). These SMs, in turn, help plants mediate both abiotic and biotic stressors such as UV light and herbivory. In the case of the betalains, there is some evidence that these SMs have some advantages over their non-tyrosine derived functional homologs in other plants, such as being more pH stable, and therefore providing more protection for CAM plants than the more common anthocyanin pigments (Jain and Gould, 2015).

Conversely, arogenate is also the precursor to the amino acid phenylalanine (Phe), which is itself a major precursor to phenylpropanoid SMs, such as flavonoids, lignins, and anthocyanins. These SMs are canonically a huge component of specialized metabolism, aiding in structure, pollinator attraction, and defense. For example, anthocyanin pigments, which range from red to blue, play many roles, including photoprotection, pollinator attraction, and antioxidant action (Landis et al., 2015). However, because Tyr and Phe are competing for the same precursor, the duplication and relaxed regulation of ADH in Caryophyllales means less arogenate is being fed into the Phe synthesis and the downstream pathways.

One apparent consequence of this competition between Tyr- and Phe-dominant SM pathways is the mutual exclusivity of betalains and anthocyanins in the order. In plants, betalains are exclusively found in Caryophyllales. However, the anthocyanins, derived from Phe, are ubiquitous throughout the plant kingdom. In lineages that have betalain, anthocyanin is never

found, and vice versa (Clement and Mabry, 1996). Although the gain of ADH-α appears to have occurred only once, betalain itself likely had multiple origins in Caryophyllales, leading to a pattern of anthocyanin-producing families sister to betalain producing ones (Sheehan et al., 2020; Fig. 2.1). These later-diverging anthocyanin-producing lineages (the most notable being the order's namesake, Caryophyllaceae) tend to lose the function of their ADH-α or lose it entirely, making the type of pigment a species produces a good proxy for presence/absence of ADH-α and therefore also a shorthand for Phe- vs Tyr-dominant metabolism types (Lopez-Nieves et al., 2018).

This unusual tension between SM pathways makes Caryophyllales a compelling system to study the relationship between specialized metabolism and human medicinal use. We predicted that within Caryophyllales, lineages with different dominant metabolic pathways would be associated with different categories of medicinal uses. For example, several Tyr-derived SMs such as salidroside, mescaline, and catecholamines affect the nervous system, while there is some evidence supporting anti-cancer, cardiovascular, and dermatological benefits of Phe-derived SMs (Berman et al., 2017; Casarini et al., 2020; Vamvakopoulou et al., 2023, Zhong et al., 2018). Tyrosine- and phenylalanine-derived metabolites in general typically make up a large part of SM in plants, with the phenylalanine-based phenylpropanoid pathway playing a large role in plant stress management and in potential medicinal effect for humans (Sun and Shahrajabian 2023).

To test the hypothesis that lineages with Tyr- and Phe-dominant specialized metabolism have different categories of use in traditional medicine, we used a phylogenetic hot nodes approach to explore over- and under-representation of different categories of medicinal use in selected lineages of the order, with representatives from both betalain and anthocyanin-

producing families (Pellicer et al., 2018; Saslis-Lagoudakis et al., 2012; Zaman et al., 2021). To test whether a factor other than metabolism, such as proximity to humans, may be driving medicinal plant selection, we used Global Biodiversity Information Facility (GBIF) occurrence data on all North American members of the Caryophyllales to explore the relationship between medicinal use and how apparent or frequently encountered and noticed a species is to humans via phylogenetic generalized least squares regression (PGLS).

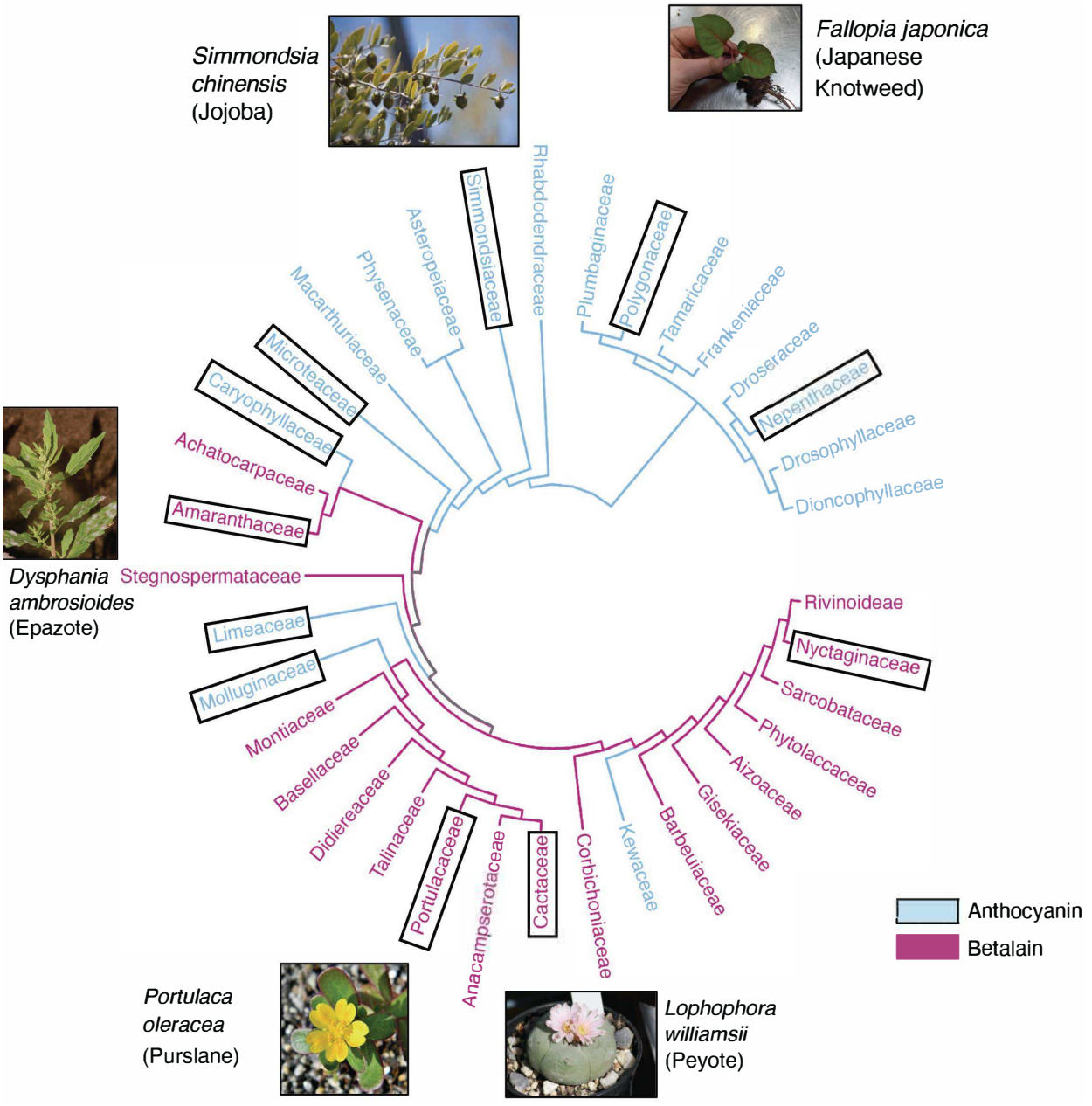


Fig. 2.1-A family level phylogeny of Caryophyllales. Blue branches represent anthocyanin- producing lineages, while magenta represents betalain-producing ones. Families surveyed in the global medicinal use literature review are outlined in black. Representatives of well-known medicinal species are shown next to the phylogeny, including *Simmondsia chinensis* (Simmondsiaceae), *Fallopia japonica* (Polygonaceae), *Dysphania ambrosioides* (Amaranthaceae), *Portulaca oleracea* (Portulacaceae), and *Lophophora williamsii* (Cactaceae). Photo credits: Ken Bosma (Jojoba); Barry Hammel (Epazote); Bob Peterson (Purslane); CC-BY-2.0

METHODS

Literature Search—At ~12,500 species, assembling all known medicinal uses in Caryophyllales is prohibitive. Therefore, we selected families that represent the phylogenetic and metabolic diversity of the Caryophyllales (Fig. 2.1). These include ancestrally Phe-dominant, anthocyanin-producing families (Polygonaceae, Nepenthaceae, and Simmondsiaceae), Tyr-dominant betalain-producing families (Cactaceae, Portulacaceae, and Amaranthaceae), and anthocyanin-producing families in later-diverging Caryophyllales representing loss of ADH and reversals to Phe-enriched metabolism (Caryophyllaceae, Limeaceae, Molluginaceae, and Microteaceae; Lopez-Nieves et al. 2018). Apart from representing either Phe- or Tyr-dominant specialized metabolism using contrasting pigment types as a proxy, these families either have well-known species with medicinal use (i.e. San Pedro cactus, rhubarb, and chickweed), or are of manageable sizes. The selection of families also ensure that the tree is roughly balanced between anthocyanin- and betalain-producing lineages. Additionally, to mitigate the potential of a specific climate selecting for metabolites of a given effect, these families represent a range of climatic conditions, with desert, tropical, and temperate species represented in selected anthocyanin and betalain producing families with an overall global distribution.

A review of the medicinal uses in these families across the globe was conducted using Google Scholar, the Native American Ethnobotany Database (NAEB), and the PubMed database

via the R package rentrez (Moerman, 2003; Winter, 2017). When taxonomy in the reported medicinal use differed from the species name on the guide tree (see methods below), it was checked using Plants of the World and Tropicos and reconciled (POWO, 2023; Tropicos). Specific medicinal use and culture/region were recorded for each species. Medicinal uses were then categorized into broader categories based on physiological system (e.g. respiratory, cardiovascular, or digestive) according to Level 2 of the Economic Botany Data Collection Standard (Cook, 1995). The region of reported use for a plant species was categorized using the M49 standard of the United Nations Statistics Division.

***Testing for phylogenetic clustering*—**Phylogenetic clustering of medicinal use was tested using a "hot nodes" approach, a method originally designed to test for phylogenetic clustering in an ecological community (Ernst et al., 2016; Saslis-Lagoudakis et al., 2012). In this method, nodes on a phylogeny with statistical overrepresentation of species with a reported medicinal use in a given category are "hot". Additionally, nodes with statistical underrepresentation are "cold" nodes. The Caryophyllales phylogeny as published in Smith et al. (2018) was trimmed to the selected families using the ape package in R, resulting in a tree with 3629 tips (Paradis and Schliep, 2019). Over- and under-representation of each category of medicinal use across this guide tree was tested using the "nodesig" command in Phylocom v4.2 (Webb et al., 2008). This function takes the number of species in a medicinal category and simulates a category with the same number of species randomized across the tree. For each medicinal category, we used the default number of 999 simulated runs. If a node includes more medicinal species of that category than that node in 97.5 % of simulated runs, it is classified as a hot node. If it has fewer medicinal species than 97.5% of the simulations, it is classified as a cold node. Therefore, each node was

classified as hot, cold, or insignificant for each medicinal use category with no associated confidence interval. A hot nodes analysis was run for each medicinal category in the global analysis of selected families as well as all Caryophyllales families in North America. At the global level, due to the low number of reported uses in the Nervous System and Mental Disorders categories and their action on the same body system, the two categories were combined for hot nodes analysis. All other categories were kept separate. Phylocom output, region of use, and reported species were visualized on the phylogeny in R using the ggtree and ggtreeExtra packages (Yu et al., 2017; Xu et al., 2021).

Phylogenetic regression of North American medicinal use—To explore the relationship between medicinal use and apparency to humans, medicinal uses exclusively from North America were collected and tested for hot nodes as described above for the entirety of Caryophyllales. Additionally, occurrence records of all Caryophyllales species from the M49‘s standard of North America (Bermuda, Greenland, the USA, Canada, and St. Pierre and Miquelon) were downloaded from GBIF, cleaned, and quality checked using the R package CoordinateCleaner (United Nations Statistics Division, 1999; Zizka et al., 2019). Dataset cleaning included removing datapoints with zero or without latitude and longitude, without a country code corresponding to the M49 standards definition of North American countries, and those with locations that matched a country centroid, administrative capital, or biodiversity institution (United Nations Statistics Division, 1999). Datapoints without a species name were also dropped. For filtering by institution or capital, datapoints within a radius of 1000 m around the site’s latitude and longitude was removed. Occurrences belonging to species not on the guide tree were discarded. Reciprocally, species not in the occurrence dataset were trimmed from the

guide tree. Each remaining species on the guide tree was classified as “medicinal” or “not medicinal”. Occurrence records for each species were counted and used directly in PGLS as a proxy for apparency, which was tested for correlation with medicinal species. The PGLS was conducted in R using the package caper (Orme et al., 2012).

Analysis was constrained to one region to mitigate any potential bias due to some regions having a more complete ethnobotanical record in literature than others. North America is an ideal choice because, for one, it has a relatively comprehensive ethnobotanical record, thanks in part to NAEB. However, the general population is, for the most part, not reliant on traditional plant medicine and there is less emphasis on native medicinal plant research here than in regions such as southern and southeastern Asia (Moerman, 2003). This means that in North America species are less likely to be collected or observed specifically because they are medicinal, and the number of occurrence records can be a good proxy for apparency.

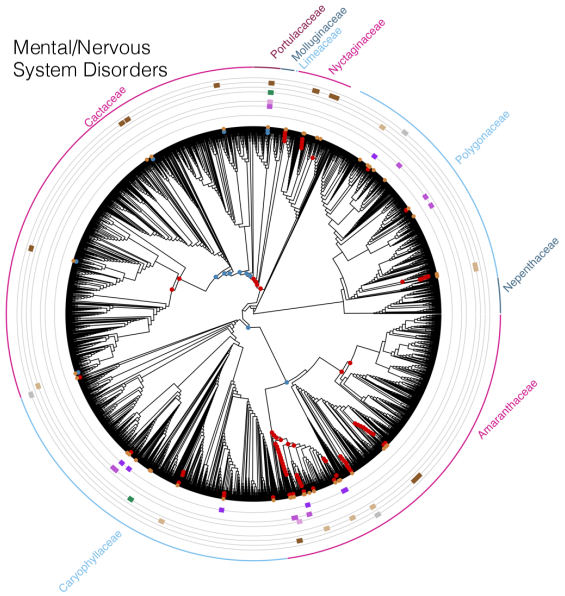
RESULTS

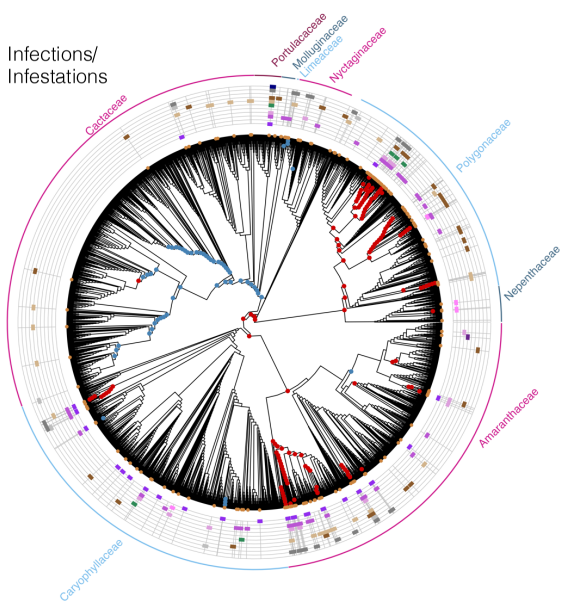
***Global analysis of selected Caryophyllales families*—**A total of 1781 categorized medicinal uses matching 465 species across 11 Caryophyllales families were retrieved in the literature search. Of those, 319 species (69%) with 1365 uses overlapped with the guide tree. The species not on the guide tree were discarded from further analysis.

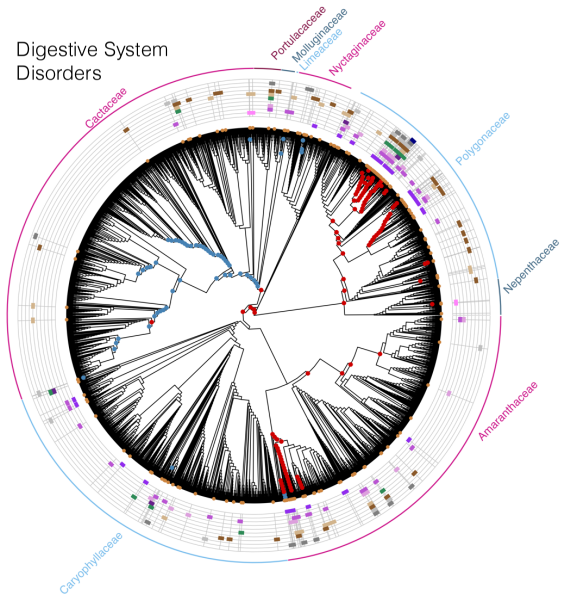
The most uses occurred in the Digestive System Disorders category (Table S2.1). North America had the most species with reported uses (108 species); however, the most reported uses across species and categories come from Southern Asia (392 uses;, Table S2.2). Eastern Asia had the highest average number of different use categories per species (4.28; Table S2.2).

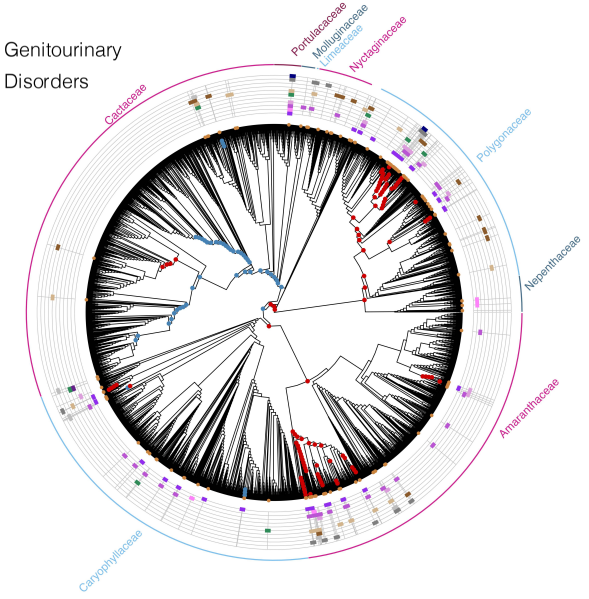
Clades with hot nodes were largely consistent among categories, although categories with fewer reported uses predictably had fewer hot nodes (Figure 2.2; Figure S2.1). Cactaceae, a

betalain-producing family with 1108 species represented on the guide tree (out of approximately 1851 total species in the family; 60%), consistently had scattered cold nodes and only a few, mostly deep, hot nodes across the categories, with the notable exception of the Endocrine System Disorder category, which had a hot clade in the *Opuntia* genus of Cactaceae (Korotkova et al., 2021; Table S2.1, Table 2.3; Fig. S2.1). Outside of Cactaceae, cold nodes were relatively uncommon. On the other end of the spectrum, the anthocyanin-producing family Polygonaceae, with only 565 of about 1200 species (47%) represented in the guide tree, was fairly consistently overrepresented across medicinal use categories (Burke and Sanchez, 2011; Fig. S2.1**;** Fig. 2.2). Hot nodes in Polygonaceae mostly occurred in the clades representing the Rumiceae and Persicarieae. Hot nodes in Caryophyllaceae were few and less consistent across categories, with medicinal uses scattered across the family (Fig. 2.2). Amaranthaceae hot nodes mostly occurred in the Amaranthoideae and Gomphrenoideae subfamilies, with less consistent hot nodes in the Chenopodioideae subfamily occurring in some categories (Fig. 2.2; Fig. S2.1).


Mental/Nervous
System Disorders
Portulacaceae
Molluginaceae
Limeaceae
Nyctaginaceae
Cactaceae
Polygonaceae
Nepenthaceae
Amaranthaceae
Caryophyllaceae


Infections/
Infestations
Portulacaceae
Molluginaceae
Limeaceae
Nyctaginaceae
Cactaceae
Polygonaceae
Nepenthaceae
Amaranthaceae
Caryophyllaceae


Digestive System
Disorders
Portulacaceae
Molluginaceae
Limeaceae
Nyctaginaceae
Cactaceae
Polygonaceae
Nepenthaceae
Amaranthaceae
Caryophyllaceae


Genitourinary
Disorders
Portulacaceae
Molluginaceae
Limeaceae
Nyctaginaceae
Cactaceae
Polygonaceae
Nepenthaceae
Amaranthaceae
Caryophyllaceae


Region
Eastern Asia
Southeastern Asia
Southern Asia
Western Asia
Central Asia
Europe
Latin America & Caribbean
North America
Northern Africa
Sub-Saharan Africa
Oceania
Cold Node
Hot Node
Medicinal Species

Fig. 2-Medicinal use in four major categories in representative families of Caryophyllales across the globe. Red nodes indicate clades where medicinal use is phylogenetically clustered in that category. Gray nodes indicate clades where medicinal use is underrepresented. The geographical region(s) where a medicinal species is used in that category is given in concentric rings outside the phylogeny. The outermost ring of magenta and blue indicate clades that are either anthocyanin or betalain producing. The top three medicinal categories are shown here, as well as the Mental/Nervous System Category, to illustrate its pattern of little recorded use as discussed in the text, and potential tyrosine-specific pattern. However, the other, less abundant medicinal categories show the same pattern of hot nodes in Polygonaceae and Amaranthaceae and cold nodes in Cactaceae (with the exception of Endocrine System Disorders; Fig S2.1).

***North American analysis of all Caryophyllales families*—**Based on the overlap with North American Caryophyllales GBIF data, the guide tree was trimmed to 914 species, 103 of which had medicinal use in at least one category, for a total of 300 medicinal uses. Similar patterns of use were observed, although hot clades were less consistently present across categories due to the reduced sample size. When present in Amaranthaceae and Polygonaceae, hot nodes were in the same clades highlighted in the global analysis. Nyctaginaceae, now a larger percentage of the tree, also had hot nodes in several categories on the North American tree. Cold nodes overall were less common in the North American analysis, although Cactaceae still had few, inconsistent hot nodes across categories, again with the exception of the "Endocrine System Disorders" category (Fig. 2.3; Fig. S2.2**)**. Medicinal use significantly correlated with a higher level of apparency ($F = 171.2$, $df = 1$, $P<2.2e\text{-}16$, $\lambda = 0.005$; Fig. 2.4).

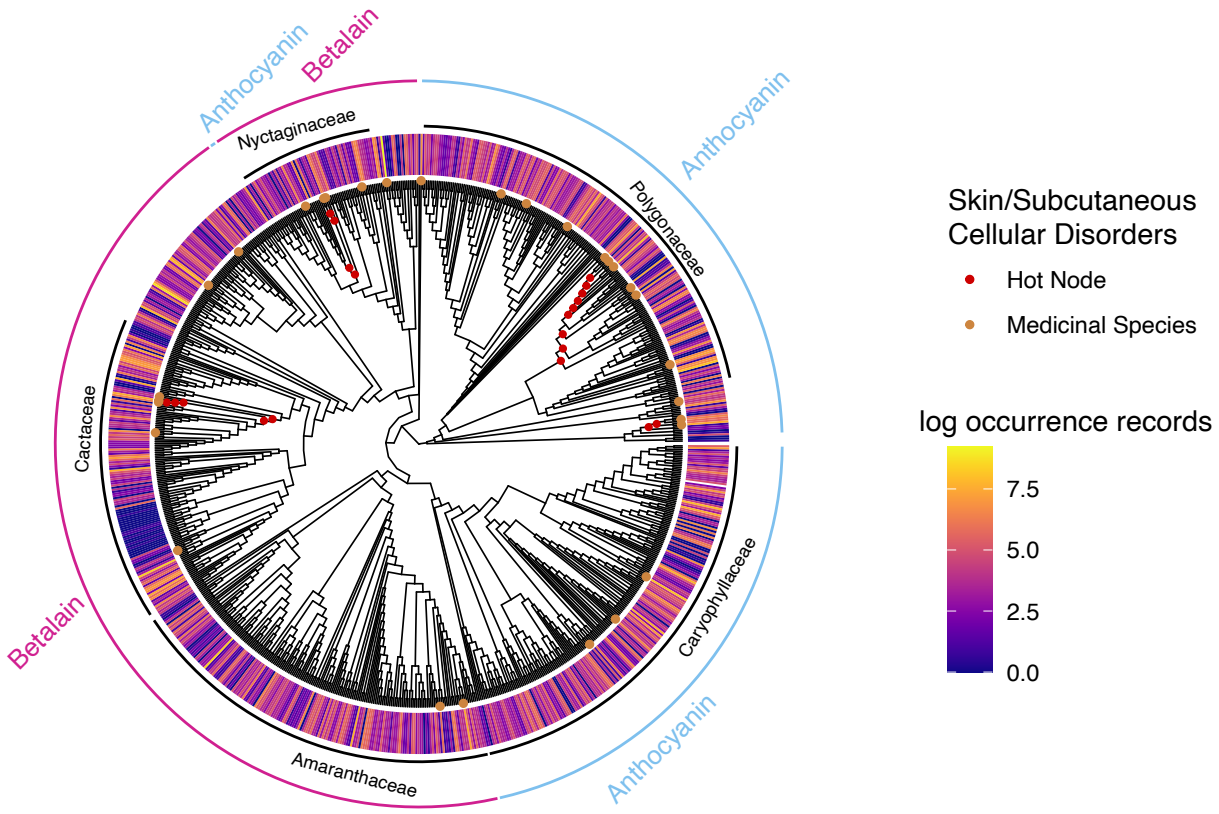


Fig. 2.3- A hot nodes analysis of Skin/Subcutaneous Cellular Tissue Disorders' medicinal use in the Caryophyllales species of North America. The ring outside the phylogeny represents log-transformed occurrences from GBIF.

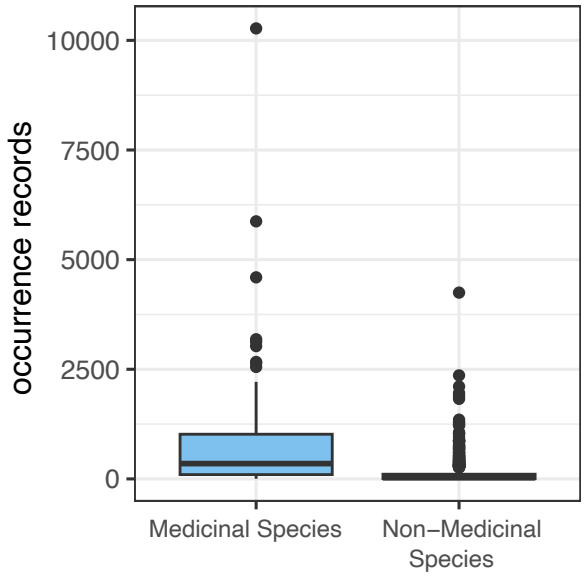


Fig. 2.4- Boxplot of number of occurrence data from GBIF of North American Caryophyllales species with any type of medicinal use vs. those with no reported medicinal use. The numbers of occurrence records on GBIF are plotted on the y-axis, with the assumption that a greater number of records represents greater apparency. Species were categorized as either medicinal or non-medicinal.

DISCUSSION

Despite the dichotomy of Tyr or Phe-dominant metabolism in Caryophyllales, overall, reported traditional medicinal uses did not differ by metabolism type. Instead, the same clades were largely selected for medicinal use repeatedly across all categories and geographic regions. In the global analysis, the two main hot clades in Polygonaceae and Amaranthaceae were also notable for having a pattern of use across many regions in all categories.

Species in these hot clades also tend to be common or even weedy and have a wider geographic range of distribution. For example, the Polygonaceae genus *Rumex*, which is consistently overrepresented across medicinal use categories, is worldwide in distribution, with several species considered invasive in the U.S. In the larger categories, such as Digestive System Disorders, most *Rumex* species are also used in multiple regions. *Rumex* species used in multiple categories are also usually used in multiple regions depending on category (i.e. the ubiquitous *Rumex crispus* L.). On the other hand, Cactaceae, which had few hot nodes, has many slow-growing species largely restricted to North and South America. Although a few *Opuntia* species

have been distributed worldwide by humans and are used outside their native range, for the most part, Cactaceae's medicinal use is restricted to the North America and Latin America/Caribbean regions. Common species being disproportionately used medicinally has been observed in several cultures and locations previously in the Highland Maya of Mexico and South Africa outside of this work as well (Lewu and Afolayan, 2009; Stepp and Moerman, 2001).

The North American analysis supports apparency associated with medicinal plant selection—The pattern of higher medicinal use in common, widespread species from our global analysis suggests that plant species that have more contact with humans and a variety of human cultures are more likely to be selected for medicinal use across categories than the species less apparent to humans. Here, the apparency of a plant species is a measure of how common a plant species is and how easy it is for humans to notice and interact with it. Higher apparency plants being used more medicinally is supported by our additional analysis restricted to North America but including all Caryophyllales families.

Because North America has less of a medicinal plant research emphasis than countries like India and China, GBIF data from North America has less bias towards medicinal plants being observed specifically because they are medicinal. Therefore, the North American analysis allows us to use occurrence data as a proxy for apparency without introducing a bias of species being observed specifically because they are medicinal. Additionally, with its simultaneous expansion of families considered and restriction of species to a specific region, the North American analysis supports the patterns of medicinal use observed in the global analysis being grounded in actual patterns of human use, rather than geographic or taxonomic selection bias. While hot nodes are less consistent across medicinal use categories in North America, which

isn't surprising given the reduced sample size of species on the tree and used medicinally, the same clades in Polygonaceae and Amaranthaceae tend to either have hot nodes or be one of the few clades with medicinal use of any kind. Nyctaginaceae, now a larger proportion of the tree than in the global analysis, still has relatively few and scattered hot nodes. The preference of certain clades across North American cultures, the attribution of many medicinal uses across categories to the same clades, and the significant correlation between occurrence data and whether or not a plant is used medicinally supports our hypothesis that in Caryophyllales, apparency is strongly associated with medicinal plant selection.

***The availability and resource availability hypotheses*—**Apparency being strongly associated with medicinal use is consistent with the availability hypothesis that accessible species, often growing in disturbed areas, are more likely to be selected for medicinal use due to frequent human contact (Gaoue et al., 2017; Stepp and Moerman, 2001; Voeks, 2004). Some of these species spread with humans as they move across the globe, increasing the likelihood of contact with different cultural groups, and helping them acquire uses in new categories (Chapman et al., 2017). The availability hypothesis also explains the lack of use in Cactaceae, which is mostly restricted to North and South America, with the exception of several widely introduced species (i.e. *Opuntia* spp.). Tellingly, the species of Cactaceae used in the most categories (15) is *Opuntia ficus-indica,* which also sees use in the most regions (6). Additionally, as Stepp and Moerman (2001) point out, humans are more likely to seek out medicines that are abundant and easily acquired when necessary, rather than relying on uncommon or difficult-to-access plants. While the availability hypothesis argues that underlying metabolism is not the

primary factor underlying human selection for medicinal use, there may be other factors to consider.

There is prior evidence to suggest that faster-growing and shorter-lived plant species tend to have more quickly made, bioactive, and therefore potentially medicinal, compounds rather than investing in longer-lasting, but slower and costly to produce defenses such as lignins and tannins (Endara and Coley, 2011). This is in line with the resource availability hypothesis, which posits that plants adapted to high-resource areas, such as human-disturbed areas, tend to be quick-growing and therefore put more resources into "quick" herbivory defenses, like bioactive specialized metabolites, rather than "slow" structural defenses that reduce the plant's digestibility (Endara and Coley, 2011; Gaoue et al., 2017). In the context of this study, the resource availability hypothesis suggests the more apparent species get used because they truly do tend to produce more medicinal compounds than their relatives growing in more resource-scarce areas. This would also explain the relative lack of medicinal uses in Cactaceae, a comparatively slow-growing group that may invest more in structural defenses, such as spines and glochids, than bioactive SMs. The patterns of medicinal use may also be due to a combination of both apparency and SM, where apparent species have evolved to have more SMs to deter herbivory, and also are more available to be selected for medicinal use. It should also be noted that there are examples outside of Caryophyllales of plants with limited apparency or a slow growth rate having significant medicinal effects, such as the tropical montane trees of the *Cinchona* genus (González-Orozco et al., 2023). Therefore, further exploration of the relationship among medicinal use, apparency, and herbivory defense in Caryophyllales is needed.

While statistical power is limited in medicinal categories with few uses, and it is difficult to separate the two potential drivers of selection, there may still be some signal of metabolite-

based selection under the overwhelming signal of apparency-driven selection. It is compelling to note that global analysis of the Mental/Nervous System Disorders category has few uses in Polygonaceae (an anthocyanin-producing family), and the only two hot nodes in the family occur at the tips (Figure 2.2). Although a few of the other small categories (i.e. Sensory System Disorders and Pregnancy/Birth/Puerperium Disorders) also lack hot nodes in Polygonaceae, they typically also show little clustering in Amaranthaceae (Fig. S2.3). In the Mental/Nervous System Disorders Category, the Amaranthaceae (a betalain lineage) hot nodes still show clustering. Further, in the North American analysis of Mental and Nervous System Disorders, there is not a single medicinal use in anthocyanin lineages, with most occurring in either Nyctaginaceae, Amaranthaceae, or scattered through Cactaceae. Mental/Nervous System disorders may be generally underrepresented in the medicinal plant toolchest, as Halse-Gramkow et al. (2016), found just over 500 psychoactive species in a review of the entire plant kingdom. Although their literature search was based primarily on a few encyclopedias of known psychoactive plants, and like our literature search here, is not fully comprehensive, 501 species across the plant kingdom is notably small when considering that in just the families selected for our global analysis, there are 165 species used in the Infections/Infestations category. Despite limited uses reported for these categories, the even more limited use in anthocyanin lineages suggests the predicted pattern of more use in lineages with Tyr-dominant metabolism. Halse-Gramkow et al. (2016) also recovered hot nodes in Cactaceae in their analysis, although they did not discuss the rest of Caryophyllales. A more targeted exploration of use of Caryophyllales species in this category is needed to definitively distinguish this pattern from statistical noise and the potential bias of there being more betalain species than anthocyanin species on the global and North American trees in this study.

***Caveats*—**The ethnobotanical record available in the literature is likely incomplete, and a substantial body of work does not exist on PubMed, Google Scholar, and our queried databases, such as non-digitized books. This is evident from the somewhat biased geographic patterns observed in our data. The most reported uses we recovered come from Eastern Asia, where Traditional Chinese Medicine has been actively studied and documented; whereas the most medicinal species come from North America, which has an extensive database of traditional use (Moerman, 2003). It is possible that we might start to see differences emerge in medicinal categories if more ethnobotanical research was available using our search approach in less studied areas such as Oceania, which has native and introduced species from several of the families reviewed here, but only three medicinal species (*Rumex maderensis* Lowe*, Portulaca oleracea* L.*,* and *Amaranthus viridis* L.*)* recovered in our literature search. However, the fact that the hot nodes we recovered are often at the base of clades with wide geographic distributions suggests that, with additional documentation of medicinal use in some regions, these nodes would still be consistently hot across categories. Additionally, the North American analysis, while having less consistent hot nodes across medicinal use categories, did not show a radically different pattern of where the hot nodes were placed when present compared to the global analysis.

The correlation between a SM and medicinal use is also likely not a linear one. Potential interactions of multiple plants in a medicinal recipe or the use of one plant to boost another's effect cannot be discounted, as this could contribute to a medicinal effect in a way individual species would not be reported in literature. For example, interactions between *Psychotria viridis* Ruiz & Pav. (Rubiaceae) and *Banisteriopsis caapi* (Spruce ex Griseb.) C.V. Morton

(Malpighiaceae) in a common Ayahuasca recipe led to its psychoactive effect. The active compound from *P. viridis* would not be able to reach the central nervous system without SMs from *B. caapi* protecting it from degradation, a relationship that has only been understood after extensive investigation of these plants' SMs (Gambelunghe et al., 2008). It is also possible that not all plant species in the indigenous pharmacopeia have a pharmacological effect matching their reported uses. Evidence of the efficacy of some common herbal medicines can vary by study, with some studies finding little to no effect for their reported uses (Bent and Ko 2004; Lin et al. 2022).

Another caveat is phylogenetic uncertainty in the guide tree used, especially deep relationships in Cactaceae (Fig, S2.3). Phylocom's nodesig command is unable to account for either branch length or phylogenetic uncertainty, so hot node estimation may be biased by a poorly supported tree. However, given the overall limited medicinal use reported in Cactaceae and that 38.7% of the family's medicinal species occur in one genus, *Opuntia*, the lack of hot nodes in the family is unlikely to be an artifact from phylogenetic uncertainty in the family.

Additionally, occurrence data from GBIF is an imperfect measure of plant apparency. The likelihood of humans encountering and noticing a plant is difficult to measure directly, but occurrence data by its very nature serves as a proxy, as it requires a human to interact with the plant in question. In fact, one of the well-known biases in GBIF and occurrence data in general is a bias towards disturbed or urban areas, while remote locations remain under-collected (Bowler et al., 2022; Petersen et al., 2021). While being an issue in some other uses of occurrence data, in our approach, this bias works in our favor to indicate apparency. However, geographical, temporal, and taxonomic biases also exist in GBIF, and while also probably correlated with apparency, they can also be driven by other factors, such as the research interests of data

contributors (Meyer et al. 2016). Additionally, GBIF data, as a massive dataset with many contributors, is known to have curation issues (Zizka et al. 2020). While we did take steps to clean questionable datapoints from our analysis, datapoints that were misidentified or incorrectly georeferenced may remain. However, as our approach for estimating appearance relies on counts of records within a chosen region, we do not require as high-quality georeferencing as some other uses for GBIF data, such as species distribution modeling, which requires precise location data.

Apart from methodological caveats, the weak association between medicinal use category and the dominant metabolism type may alternatively be driven by overlap between the bioactivity of the metabolites derived from the two related amino acids. Phenylalanine-derived specialized metabolism includes phenolics like flavonoids, coumarins, and lignans**,** which have been ascribed a wide range of potential health benefits. While tyrosine-derived metabolites tend to be less ubiquitous, some of the best-known include neurologically active compounds like mescaline, dopamine, and morphine (Xu et al., 2019). However, other Tyr-derived compounds may have as broad action as Phe metabolism. For example, apart from neurological effect, some Tyr-derived catecholamines like epinephrine act on the cardiovascular and respiratory system (Abul-Ainine, 2002; Bao et al., 2007). If Tyr- and Phe-derived metabolism overall have enough overlap in medicinal activity, human selection of apparent plants would be mainly driven by aparency and less by the dominant metabolism type.

Alternatively, metabolites derived from precursors other than Tyr or Phe may drive broad patterns of medicinal activity. At least one species in Caryophyllales is best known for specialized metabolites from a different pathway: *Simmondsia chinensis*, used largely in the Skin/Subcutaneous Cellular Tissue Disorders category is famous for the lipids in its oil.

However, given the overall importance of the phenylalanine-derived phenylpropanoid pathway in plant specialized metabolism across the plant kingdom, broad-scale differences in medicinal use between lineages would still be expected to be affected by the switches between Tyr- and Phe-dominant metabolism.

CONCLUSIONS

Given the renewed interest in natural products for drug discovery, understanding the relationships among specialized metabolism, phylogeny, and medicinal use is important in prioritizing plants to investigate and narrow down the metabolites responsible for medicinal function. In addition, an approach combining phylogenetic, metabolic, and global use information can elucidate cross-cultural patterns of medicinal use and what is driving them. Here, we show that, at least in Caryophyllales, the apparency of a plant to humans is a dominant factor in its being selected for medicinal use across cultures, as opposed to the plant's broad SM profile. Although there are several possible explanations for this, a combination of humans selecting readily apparent plants (availability hypothesis) and those same plants having greater bioactive activity (resource availability hypothesis) appear to be the most likely. Further investigation of the Tyr- and Phe- derived metabolic diversity in both medicinal and non-medicinal clades of the order would help determine which or both of these hypotheses is driving selection of medicinal plants in Caryophyllales and help untangle the effects of apparency from other potential drivers of medicinal plant selection.

DATA AVAILABILITY STATEMENT

Data and scripts used in this analysis are deposited on Dryad (https://doi.org/10.5061/dryad.6q573n64z)

SUPPLEMENTAL MATERIALS

Table S2.1 –Table with medicinal use counts by category, region, and family

Figure S2.1 – Hot and cold nodes mapped out on a phylogeny of selected Caryophyllales families across the globe. All medicinal categories not part of Fig. 2.2 are shown. Color of family names follow Fig. 2.1.

Figure S2.2 – Hot and cold nodes mapped out on a phylogeny of North American Caryophyllales species. A selection of medicinal categories is included.

Figure S2.3 – Guide tree used in the global analysis with bootstrap support plotted at each node.

LITERATURE CITED

Abul-Ainine, A., D. Luyt, 2002. Short term effects of adrenaline in bronchiolitis: a randomized controlled trial. *Archives of Disease in Childhood* 86: 276-279.

Afendi, F.M, T. Okada, M. Yamakazi, A. Hirai-Morita, Y. Nakamura, K, Nakamura, S. Ikeda, H. Takahashi, M. Altaf-Ul-Amin, L.K, Darusman, K. Saito, S. Kanaya. 2012. KNApSAcK Family Databases: Integrated Metabolite—Plant species Databases for Multifaceted Plant Research. *Plant and Cell Physiology* 53: e1.

Atanasov, A.G., S.B. Zotchev, V.M. Dirsch, I.N.P.S.T, C.T. Supuran. 2021. Natural products in drug discovery: advances and opportunities. *Nature Revies Drug Discovery* 20: 200-216.

Bao, X., C.M. Lu, F. Liu, Y. Gu, N.D. Dalton, B.Q. Zhu, E. Foster, et al. 2007. Epinephrine is required for normal cardiovascular responses to stress in the phenylethanolamine N-methyltransferase knockout mouse. *Circulation* 116: 1024-1031.

Barros, J., J.C. Serrani-Yarce, F. Chen, D. Baxter, B.J. Venables, R.A. Dixon. 2016. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants* 2: 16050.

Bent, S., R. Ko. 2004. Commonly used herbal medicines in the United States: a review. *The*

American Journal of Medicine 116: 478-485.

Berman, A.Y., R.A. Motechin, M.Y. Wiesenfeld, M.K. Holz. 2017. The therapeutic potential of resveratrol: a review of clinical trials. *Npj Precision Oncology* 1: 35.

Bowler, D.E., C.T. Callaghan, N. Bhandari, K. Henle, M.B. Barth, C. Koppitz, R. Klenke, M. Winter, F. Jansen, H. Bruelheide, A. Bonn. 2022. Temporal trends in the spatial bias of species occurrence records. *Ecography* 2022: e06219.

Burke, J.M., A. Sanchez. 2011. Revised subfamily classification for Polygonaceae, with a tribal classification for Eriogonoideae. *Brittonia* 63: 510-520.

Casarini, T.P.A., L.A. Frank, A.R. Pohlmann, S.S. Guterres. 2020. Dermatological applications of the flavonoid phloretin. *European Journal of Pharmacology* 889: 173593.

Chapman, D., B.V. Purse, H.E. Roy, J.M. Bullock. 2017. Global trade networks determine the distribution of invasive non-native species. *Global Ecology and Biogeography* 26: 907-917.

Chen, J., Y-P. Shi, J-Y. Liu. 2003. Determination of noradrenaline and dopamine in Chinese herbal extracts from *Portulaca oleracea* L. by high-performance liquid chromatography. *Journal of Chromatography A* 1003: 127-132.

Clement, J.S., T.J. Mabry. 1996. Pigment evolution in the Caryophyllales: a systematic overview. *Botanica Acta* 109: 360-367.

Cook, F.E.M. 1995. Economic botany data collection standard. Royal Botanic Gardens, Kew, Great Britain.

David, B., J.L. Wolfender, D.A. Dias. 2015. The pharmaceutical industry and natural products: historical status and new trends. *Phytochemistry Reviews* 14: 299-315.

Endara, M.J., P.D. Coley. 2011. The resource availability hypothesis revisited: A meta-analysis. *Functional Ecology* 25: 389-398.

Ernst, M., C.H. Saslis-Lagoudakis, O.M. Grace, N. Nilsson, H.T. Simonsen, J.W. Horn, N. Rønsted. 2016. Evolutionary prediction of medicinal properties in the genus *Euphorbia* L. *Scientific Reports* 6: 30531.

F.O. Cruz. 2023. Botanical Collection of Medicinal Plants, Instituto de Tecnologia em Fármacos. Website http://cbpm.fiocruz.br/index?ethnobotany.

Gambelunghe, C., K. Aroni, R. Rossi, L. Moretti, M. Bacci. 2008. Identification of *N,N*-dimethyltryptamine and **β**-carbolines in psychotropic ayahuasca beverage.

Biomedical Chromatography 22: 1056-1059.

Gaoue, O.G., K. Yessoufou, L. Mankga, F. Vodouhe. 2021. Phylogeny reveals non-random medicinal plant organ selection by local people in Benin. *Plants People Planet* 3: 710-720.

González-Orozco, C.E., E.G. Guillén, N. Cuvi. 2023. Changes of *Cinchona* distribution over the past two centuries in the northern Andes. *Royal Society Open Science* 10: 230229.

Halse-Gramkow, M., M. Ernst, N. Rønsted, R.R. Dunn, C.H. Saslis-Lagoudakis. 2016. Using evolutionary tools to search for novel psychoactive plants. *Plant Genetic Resources: Characterization and Utilization* 14: 246-255.

Hamilton, A.C. 2005. Medicinal plants, conservation and livelihoods. *Biodiversity and Conservation* 13:1477-1517.

Holeski L.M., K. Keefover-Ring, J.M. Sobel, N.J. Kooyers. 2021. Evolutionary history and ecology shape the diversity and abundance of phytochemical arsenals across monkeyflowers. *Journal of Evolutionary Biology* 34: 571-583.

Holzmeyer, L., A.K. Hartig, K. Franke, W. Brandt, A.N. Muellner-Riehl, L.A. Wessjohann, J. Schnitzler. 2020. Evaluation of plant sources for antiinfective lead compound discovery by correlating phylogenetic, spatial, and bioactivity data. *Proceedings of the National Academy of Sciences, USA* 117: 12444-12451.

Huang R., A.J. O'Donnell, J.J. Barboline, T.J. Barkman. 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proceedings of the National Academy of Sciences* 113: 10613-10618.

Jain, G., K.S. Gould. 2015. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany* 119: 48-53.

Korotkova, N., D. Aquino, S. Arias, U. Eggli, A. Franck, C. Gómez-Hinostrosa, P.C. Guerrero, et al. 2021. Cactaceae at Caryophyllales.org – a dynamic online species-level taxonomic backbone for the family. *Willdenowia* 51: 251-270.

Landi, M., M. Tattini, K.S. Gould. 2015. Multiple functional roles of anthocyanins in plant-environment interactions. *Environmental and Experimental Botany* 119: 4-17.

Lewu, F.B., A.J. Afolayan. 2009. Ethnomedicine in South Africa: The role of weedy species. *African Journal of Biotechnology* 8: 929-934.

Lin, C-H., Y_A. Lin, S-L. Chen, M-C. Hsu, C-C. Hsu. 2022. American ginseng attenuates

eccentric exercise -induced muscle damage via the modulation of lipid peroxidation and inflammatory adaptation in males. *Nutrients* 14: 78.

Lopez-Nieves, S., Y. Yang, A. Timoneda, M. Wang, T. Feng, S.A. Smith, S.F. Brockington, H.A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytologist* 217: 896-908.

Merfield, C.N. 2022. Redefining weeds for the post-herbicide era. *Weed Research* 62: 263-267.

Meyer, C., P. Weigelt, H. Kreft. 2016. Multidimensional biases, gaps, and uncertainties in global plant occurrence information. *Ecology Letters* 19: 992-1006.

Moerman, D.E. 2003. Native American Ethnobotany Database [online]. Website http://naeb.brit.org/.

Nothias L-F., M. Nothias-Esposito, R. da Silva, M. Wang, I. Protsyuk, Z. Zhang, A, Sarvepalli, et al. 2018. Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. *Journal of Natural Products* 81: 758-767.

Nothias L-F., D. Petras, R. Schmid, K. Dührkop, J. Rainer, A. Sarvepalli, I. Protsyuk, et al. 2020. Feature-based molecular networking in the GNPS analysis environment. *Nature Methods* 17: 905-908.

Orme, D., R.P. Freckleton, G.H. Thomas, T. Petzoldt, S. Fritz, N. Isaac, W. Pearse. 2023. caper: Comparative analyses of phylogenetics and evolution in R, version 1.0.2. website: https://cran.r-project.org/web/packages/caper/index.html.

Paradis, E., and K. Schliep. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.

Pellicer, J., C.H. Saslis-Lagoudakis, E. Carrió, M. Ernst, T. Garnatje, O.M. Grace, A. Gras, et al. 2018. A phylogenetic roadmap to antimalarial *Artemisia* species. *Journal of Ethnopharmacology* 225: 1-9.

Petersen, T.K., J.D.M Speed, V. Grøtan, G. Austrheim. 2021. Species data for understanding biodiversity dynamics: The what, where, and when of species occurrence data collection, *Ecological Solutions and Evidence* 2: e12048.

POWO. 2023. Plants of the world [online]. Royal Botanical Gardens, Kew. Website http://www.plantsoftheworldonline.org/.

Rodman, J.E., P.S. Soltis, D.E. Soltis, K.J. Sytsma, K.G. Karol. 1998. Parallel evolution of

glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. *American Journal of Botany* 85: 997-1006.

Saslis-Lagoudakis, C.H., V. Savolainen, E.M. Williamson, F. Forest, S.J. Wagstaff, S.R. Baral, M.F. Watson, et al. 2012. Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proceedings of the National Academy of Sciences, USA* 109: 15835-15840.

Saslis-Lagoudakis, C.H., B.B. Klitgaard, F. Forest, L. Francis, V. Savolainen, E.M. Williamson, J.A. Hawkins. 2011. The use of phylogeny to interpret cross-cultural patterns in plant use and guide medicinal plant discovery: and example from *Pterocarpus* (Leguminosae). *PLoS One*: 6: e22275.

Smith, S.A., J.W. Brown, Y. Yang, R. Bruenn, S.F. Brockington, J.F. Walker, N. Last, et al. . 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist* 217: 836-854.

Stepp, J.R., D.E. Moerman. 2001. The importance of weeds in ethnopharmacology. *Journal of Ethnopharmacology* 75: 19-23.

Sun, W.. M.H. Shahrajabian. 2023. Therapeutic potential of phenolic compounds in medicinal plants – Natural health products for human health. *Molecules* 28: 1845.

Thulin, M., A.J. Moore, H. El-Seedi, A. Larsson, P.A. Christin, E.J. Edwards. 2016. Phylogeny and generic delimitation in Molluginaceae, new pigment data in Caryophyllales and the new Corbichoniaceae. *Taxon* 65: 775-793.

United Nations Statistics Division. 1999. Standard country or area codes for statistics use. Revision 4.

Vamvakopoulou, I.A., K.A.D. Narine, I. Campbell, J.R.B. Dyck, D.J. Nutt. 2023. Mescaline: The forgotten psychedelic. *Neuropharmacology* 222: 109294.

Voeks, R.A. 2004. Disturbance pharmacopeias: medicine and myth from the humid tropics. *Annals of the Association of American Geographers* 94: 868-888.

Wink, M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 64: 3-19.

Winter, D.J. 2017. Rentrez: An R package for the NCBI eUtils API. *The R Journal* 9: 520-526.

Webb, C.O., D.D. Ackerly, S.W. Kembel. 2008. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* 24: 2098-2100.

Xu, S., Z. Dai, P. Guo, X. Fu, S. Liu, L. Zhou, W. Tang, et al. 2021. ggtreeExtra: Compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution* 38: 4039-4042.

Xu, J.J., X. Fang, C.Y. Li, L. Yang, X.Y. Chen. 2019. General and specialized tyrosine metabolism pathways in plants. *aBIOTECH* 1: 97-105.

Yang, J., Z. Li, J. Lian, G. Qi, P. Shi, J. He, Z. Hu, M. Zhang. 2020. Brassicaceae transcriptomes reveal convergent evolution of super-accumulation of sinigrin. *Communications Biology* 3: 779.

Youssef, D., R. El-Bakatoushi, A. Elframawy, L. El-Sadek, G. El Badan. 2023. Molecular phylogenetic study of flavonoids in medicinal plants: a case study of family Apiaceae. *Journal of Plant Research* 136: 305-322.

Yu, G., D.K. Smith, H. Zhu, Y. Guan, T.T. Lam. 2017. GGTree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* 8: 28-36.

Zaman, W., J. Ye, S. Saqib, Y. Liu, Z. Shan, D. Hao, Z. Chen, P. Xiao. 2021. Predicting potential medicinal plants with phylogenetic topology: Inspiration from the research of traditional Chinese medicine. *Journal of Ethnopharmacology* 281: 114515.

Zhong, Z., J. Han, J. Zhang, Q. Xiao, J. Hu, L. Chen. 2018. Pharmacological activities, mechanisms of action, and safety of salidroside in the central nervous system. *Drug Design, Development, and Therapy*. 12: 1479-1489.

Zizka, A., F.A. Carvalho, A. Calvente, M.R. Baez-Lizarazo, A. Cabral, J.F.R. Coelho, M. Colli-Silva, et al. 2020. No one-size-fits-all solution to clean GBIF. *PeerJ* 8: e9916.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C.D. Ritter, D. Edler, H. Farooq, et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* 10: 744-751.

2023. Missouri Botanical Garden. Website Tropicos.org.

CHAPTER 3: MULTI-OMICS OF A BETALAIN AND AN ANTHOCYANIN PRODUCING SPECIES IN CARYOPHYLLALES REVEALS DIFFERENT RESPONSES TO STRESS BETWEEN PATHWAYS AND SPECIES

INTRODUCTION:

The anthocyanin pigments are nearly ubiquitous throughout the seed plants (Piatkowski et al., 2020). These pigments, which are part of the larger flavonoid chemical class, produce a wide range of colors. Anthocyanins play a variety of roles in plants, including photoprotection, pollinator attraction, herbivory deterrence, and more (Jain and Gould, 2015).

In contrast, betalain pigments occur in plants only in the core clade of the order Caryophyllales (Fig 3.1). Betalain pigments produce a range of reds and yellows, and may be functional homologues to anthocyanins, and in some situations, such as hot and arid conditions, may actually perform better (Jain and Gould 2015). Interestingly, anthocyanins and betalains are mutually exclusive, meaning that no anthocyanin-producing species also makes betalain and vice versa (Clement and Mabry 1996).

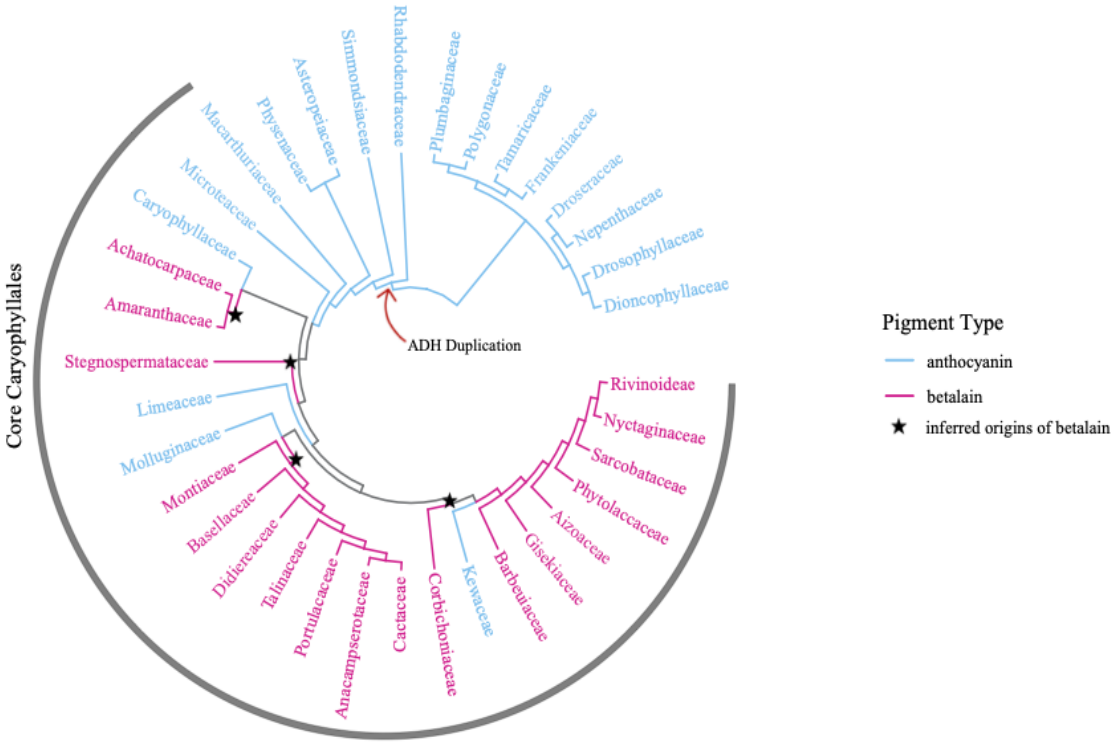


Fig. 3.1-The family level tree of Caryophyllales. The outer gray circle indicates the core clade of the order. Anthocyanin lineages are represented in blue. Betalain lineages are represented in magenta. The inferred duplication of ADH and origins of betalain are indicated. Backbone relationships and inferred ADH duplication and betalain origins are based on Lopez-Nieves et al. (2018), Sheehan et al. (2020), and Timoneda et al. (2019). Circumscription of the Core Caryophyllales follows Stevens (2014).

The development of betalains was enabled by the duplication of the gene arogenate dehydrogenase (*ADH*). This gene codes for the enzyme responsible for turning the precursor arogenate into tyrosine (Tyr). While the ancestral copy of *ADH* (*ADHβ*), is feedback inhibited, the Caryophyllales specific-copy (*ADHα*) has relaxed feedback inhibition, allowing for increased Tyr production (Lopez-Nieves et al., 2018). This availability of excess tyrosine seems to have enabled the development of the betalain pigments, which use tyrosine as a precursor. Evidence suggests that after the duplication of *ADH*, the rest of the genes for betalain synthesis were

recruited convergently in different lineages, resulting in multiple origins of betalain pigment within Caryophyllales (Sheehan et al. 2020).

The precursor arogenate, which comes out of the shikimate biosynthetic pathway, is also necessary for phenylalanine (Phe) production (Cho et al., 2007). The increase in tyrosine availability potentially comes at the cost of phenylalanine production, which may in turn contribute to the lack of anthocyanin pigments in betalain-producing lineages, as anthocyanins are synthesized from phenylalanine (Lopez-Nieves et al., 2018). However, betalain lineages retain nearly all of the main pathway genes required for anthocyanin synthesis with the exception of *TT19*, a glutathione S-transferase responsible for catalyzing one of the final steps in anthocyanin synthesis (Pucker et al., 2023).

There are a few lineages in the core Caryophyllales that never developed betalain and instead retain the ancestral anthocyanins (Caryophyllaceae, Molluginaceae,and Limeaceae), and one, Kewaceae, that may be a reversal back to anthocyanin (Sheehan et al., 2020). Within these families, *ADHα* seems to be lost, although some basal members have a copy that may or may not be functional (Lopez-Nieves et al. 2018).

Besides the pigments, tyrosine and phenylalanine are the precursors for a wide range of specialized metabolites. While tyrosine-derived metabolites (i.e. catecholamines and isoquinoline alkaloids) tend to be more lineage specific, the phenylpropanoid pathway, which converts phenylalanine to flavonoids, lignins, etc. is phylogenetically widespread, and its products tend to play very important roles across the plant kingdom, including structure and defense (Dong and Lin, 2021). In fact, the anthocyanin synthesis genes, especially the early ones, are hypothesized to be conserved in betalain lineages due to their necessity in synthesizing other phenylpropanoids (Pucker et al., 2023; Fig 3.2).

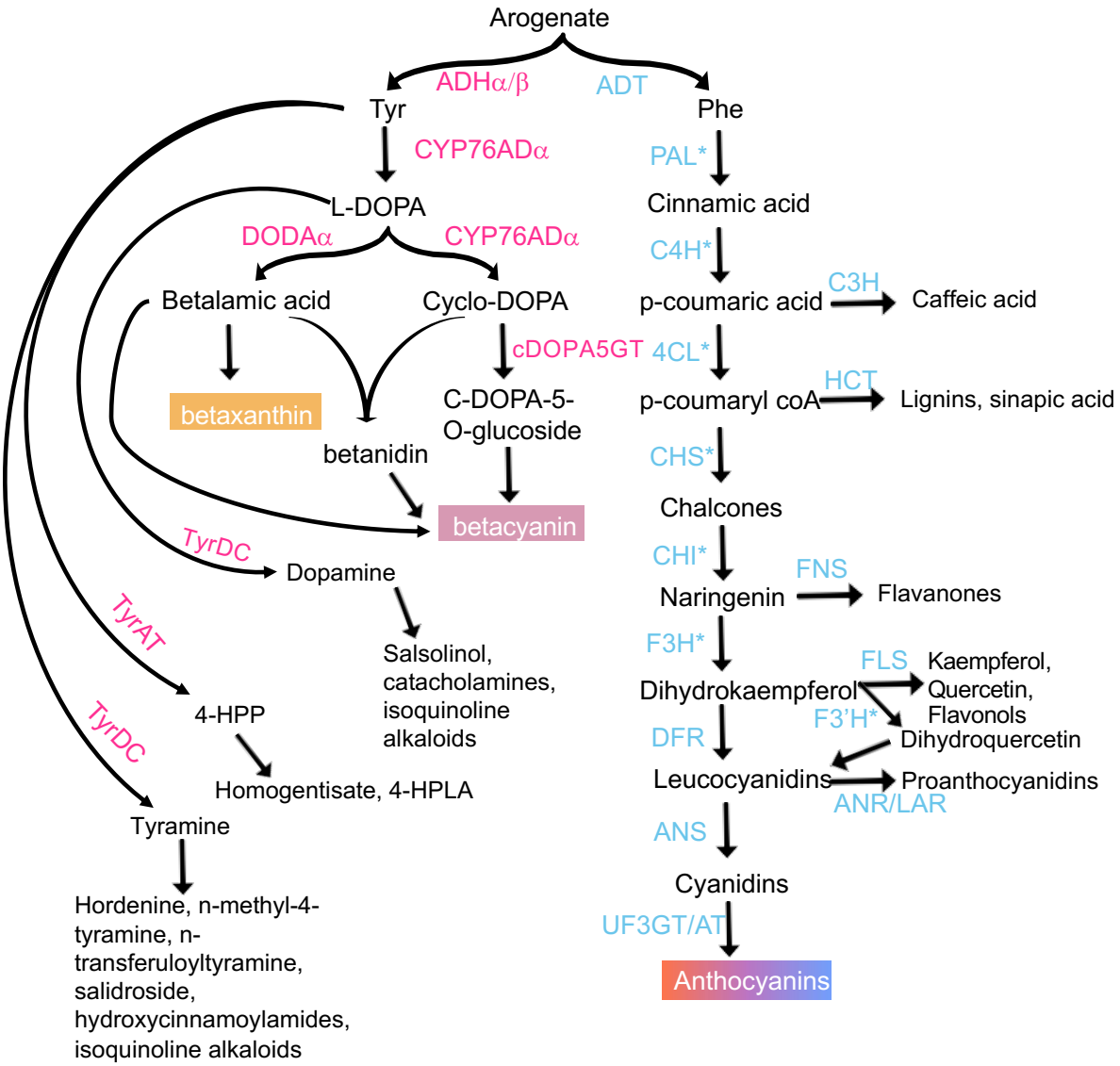


Fig. 3.2-The genes and metabolites involved in the betalain and anthocyanin synthesis pathways, plus some of the important branches off the pigment pathways. Metabolites are in black font, tyrosine-associated genes in pink, and phenylalanine genes in blue. Asterisks next to anthocyanin synthesis genes indicate that they are early biosynthetic genes.

Thanks in part to its pervasiveness and importance, the phenylpropanoid pathway is well studied in many species, particularly anthocyanin synthesis (Campanella et al. 2014; Liu et al. 2018). On the other hand, the tyrosine pathway, with the exception of betalain synthesis is less well characterized. For example, regulation of betalain synthesis is largely unknown, with the exception of a few MYB genes identified in *Beta vulgaris* L., *Amaranthus hypochondriacus* L., and *Selenicereus undatus* (Haw.) D.R. Hunt (Hatlestad et al., 2015; Winkler et al., 2024; Xie et al., 2023). Interestingly, the MYB gene in *B. vulgaris* was co-opted from the anthocyanin pathway (Hatlestad et al., 2015).

Although the wider phenylpropanoid synthesis pathway is relatively well-known in other lineages, its status post-*ADH* duplication is still poorly known. Although *ADHα* is apparently lost in the lineages that never develop betalain, it is unknown what effects the ancestral state of excess tyrosine may have had on phenylalanine metabolism in these lineages. It may be that the anthocyanin lineages in the core Caryophyllales, while having an ancestral anthocyanin state, do not have the ancestral phenylalanine metabolism. The status of phenylalanine metabolism genes in betalain-producing lineages, apart from the anthocyanin synthesis ones, is also mostly unknown. Given the functional similarities between anthocyanins and betalains, as well as the co-option of at least one anthocyanin gene, it is possible that more genes have been co-opted from phenylalanine metabolism or been functionally replaced by tyrosine pathway genes.

Because betalain and anthocyanin synthesis are two of the better characterized pathways in tyrosine and phenylalanine specialized metabolism, they are an excellent place to start in investigating the interactions of these two specialized metabolism pathways in Caryophyllales. Here, we use salt, high light, and simulated herbivory via methyl jasmonate to induce pigment production in a core anthocyanin-producing species and a betalain-producing species over a 48-

hour time course. We used a multi-omics analysis to investigate pigment synthesis and other related metabolites and genes.

Focal species

The two species compared are *Beta vulgaris* subsp. *maritima* (L.) Arcang (Amaranthaceae), member of a betalain-producing lineage, and *Silene latifolia* Poir (Caryophyllaceae), member of an anthocyanin-producing lineage. *Beta vulgaris* subsp. *maritima*, or the sea beet is a wild relative of table and sugar beets. While a betalain producer like its crop relatives, sea beet has the advantage of inducible pigment production, rather than being bred for either no or constant over-production of betalain pigmentation. Sea beet is also a halophyte, tending to grow on or near beaches or saltwater in its range, which spans north Africa to northern Europe, with introductions also occurring in North America and as far east as India (Frese and Brian Ford-Loyd, 2012).

Silene latifolia belongs to the anthocyanin producing-family Caryophyllaceae and diverged post-*ADH* duplication. However, the Caryophyllaceae likely retained ancestral anthocyanin and never developed betalain pigmentation, as evidenced from ancestral state reconstruction of L-DOPA 4,5-dioxygenase activity in *DODA* and its maintenance of *TT19* (Pucker et al., 2023; Sheehan et al., 2020). *ADHα* also seems to be lost or pseudogenized in the family (Lopez-Nieves et al., 2018; Sheehan et al., 2020). The family, therefore, despite maintaining anthocyanin pigmentation, may represent a reversal back to phenylalanine-dominant metabolism. Originally native to Europe, *S. latifolia* has become widespread as a weed in North America.

Both species, apart from representing different pigmentation states, were selected for their fast-growing annual habit and ease of germination after pre-experiments on over ten

candidate species in Caryophyllales. Both species also have reference genomes available to aid in transcriptome assembly and annotation.

We hypothesized that *Silene*, as an anthocyanin producer, has a Phe-enriched metabolism, while sea beet has a Tyr-enriched metabolism. We predicted that as a result, sea beet would express a higher diversity and abundance of Tyr-related genes and metabolites, while *Silene* expresses a higher diversity and abundance of Phe-related genes and metabolites. Additionally, we expected that the tyrosine pathway in sea beet would functionally replace the phenylalanine pathway in *Silene* and respond similarly to biotic and abiotic stressors. We found that, while sea beet's tyrosine-derived metabolites were more responsive to stress, the Phe pathway in *Silene* did not behave similarly to the Tyr pathway in sea beet. In fact, both Tyr- and Phe-derived pathways show evidence of reduced gene expression, metabolite production, and stress response in Silene. This suggests that, despite never having developed betalain and losing ADHα, Silene is not a reversal back to Phe-enriched metabolism, nor does it show evidence of Tyr-enriched metabolism, instead being an intermediate type that incorporates remnants of both.

MATERIALS AND METHODS

Seed germination and growth

We tested over 10 species across the Caryophyllales for the ease and uniformity of germination and seedling growth, and visible pigment accumulation under stress treatments. We chose *Beta vulgaris* subsp. *maritima* and *Silene latifolia* for our final experiments. Seeds for *Beta vulgaris* subsp. *maritima* were obtained from the USDA-GRIN (GRIN PI 562585), grew in the greenhouse, and selfed for one generation. Seeds for *Silene latifolia* were obtained from Andrea Berardi, originally collected from Weil am Rhein, Germany in 2020 (voucher to be submitted to

HUH). Berardi grew the wild-collected seeds, crossed a male and female, and sent seeds from 4 capsules. The sea beet time course was run in October 2021, and *Silene* in October 2022 using the same growth chambers and chamber settings. For both species, seeds were initially germinated in trays of Berger BM2 germination soil mix (Saint-Modeste, Quibec, Canada), before being transplanted to ⅜ x 5 inch Anderson Band Bottom Pots (Portland, Oregon, USA) with 66% Sungro Sunshine #1 potting soil mix (Agawam, MA, USA) and 33% perlite. Transplants were also given 1 gram 14-14-14 Osmocote fertilizer on top of the soil. Throughout seed germination and seedling growth, plants were kept in a growth chamber with 14 hours of 200 μmol/m^2/s light intensity, 70% humidity, and a daylight temperature of 23° C and night temperature of 20° C. Temperature gradually ramped up or down over 30 minutes during day/night transitions. Seedlings were watered uniformly 3x a week. Experimental treatment began when at least 124 individuals of a species had 3 fully expanded leaf pairs (in the case of *S. latifolia*) or 4 fully expanded alternate leaves (in the case of sea beet).

Treatment applications

In our pre-experiments, we tested stressors including mechanical wounding, cold, UVB, heat, and evaluated different intensity and duration of stress treatments. We also compared treatments using tissue culture vs. seedlings. We decided to use high light and salt to represent two abiotic stressors, and methyl jasmonate (MeJA) treatment to simulate herbivory. We also decided to use seedlings grown in soil to simulate natural conditions.

Within both species, plants were divided into four treatment groups, and further divided into subgroups based on timepoint sampled, for a total of 100 plants per species (Table 3.1). Control treatment plants were placed into the growth chamber at the same conditions the plants

were grown in. Plants in the methyl jasmonate (MeJA) groups were dipped into a solution of 0.5 mM MeJA (Sigma-Aldrich), 0.015% Silwet L-77 (Fisher Scientific), and 0.25% 190 proof ethanol at Time 0 (the beginning of the experiment). MeJA group plants were placed in the same growth chamber as the control group after fifteen minutes to allow volatiles to disperse. High light treatment plants were placed into a growth chamber with the same schedule as the control/MeJA growth chamber and ~1300 μmol/m^2/s light intensity throughout the experiment. Salt treatment plants were soaked for 1 hour in a 450 mM NaCl solution beginning 30 minutes before Time 0.

Table 3.1-Experimental design. Number in each cell indicates the number of biological replicates sampled for each sampling point. An empty cell indicates that no sampling or measurements were taken at that time point for that particular treatment group. In total, 100 plants were used for destructive tissue sampling and 24 plants were used for non-destructive physiological measurements. The same scheme was used for both species.

Time point	Time of sampling	Destructive tissue sampling				Non-destructive physiological measurements			
		High light	Salt	MeJA	Control	High light	Salt	MeJA	Control
0 - start	noon				4	Reuse the same 6 plants across 6 time points + 7 days post time course	Reuse the same 6 plants across 6 time points + 7 days post time course	Reuse the same 6 plants across 6 time points + 7 days post time course	Reuse the same 6 plants across 6 time points + 7 days post time course
1 - 1hr	1 pm	4	4	4	4				
2 - 3 hr	3 pm	4	4	4	4				
3 - 6 hr	6 pm	4	4	4	4				
4 - 12 hr	midnight	4	4	4	4				
5 - 24 hr	noon on Day 2	4	4	4	4				
6 - 48 hr	noon on Day 3	4	4	4	4				

Total	124	24	24	24	28	6	6	6	6

Time course

Time 0, which represented the beginning of the experiment and the start of treatments, occurred at noon, which corresponded with the midday of the growth chamber schedule. The remainder of the timepoints occurred 1, 3, 6, 12, 24, and 48 hours after Time 0, in order to capture earlier changes in gene expression and later changes in metabolite abundance. At each timepoint, except for Time 0, where only control group plants were sampled, four biological replicates of each species in each treatment group were sampled (Table 1). Time 4, which occurred at midnight, was the only timepoint to occur during the growth chamber's scheduled dark time.

Sampling of each plant consisted of flash freezing in liquid nitrogen 0.1 g leaf tissue for RNA-seq, and 0.2 g leaf tissue for metabolite analysis (between 0.05 and 0.2 g as some leaves were small). Leaf tissue was sampled from the second and third fully expanded leaves from the top. In sea beet, where leaf arrangement is alternate, the older leaf was used for metabolomics, and the younger was used for RNA-Seq. In *S. latifolia*, where leaves are opposite, no discrimination was made between the two members of the leaf pair.

Physiology measurements

In parallel with the time course with sampling, a group of twenty-four plants per species was also divided into treatment groups (6 per treatment) and kept in identical conditions to the sample plants throughout the time course. These plants were used for physiological monitoring,

to ensure that treatments were having an impact. At each time point, and one week after the final time point, the physiology plants were photographed with a ruler and color card (Fig 3.5).

We also measured quantum yield (QY), or photosynthetic capacity of the physiology group using a hand-held Fluorimeter (Fluorpen FP 110, Photon Systems, Covina, California, USA). Measurements were taken from the two leaves that would be sampled in the experimental groups. QY measurements were taken in the dark before dawn the day the experiment started, at each time point, and one week after the conclusion of the experiment to measure recovery from treatment.

Physiology group plants were also scanned using a custom-built hyperspectral camera (Middleton Spectral Vision) to measure pixel-by-pixel leaf reflectance in the 400-1,000 nm wavelength range. Spectral imaging and analysis protocols were adapted from Tirado et al. (2021). The different treatments were imaged separately, so that each scan included 6 plants. Mean spectra for each plant was used to calculate the normalized difference vegetation index (NDVI), anthocyanin reflectance index (ARI), and the photochemical reflectance index (PRI). These indices were then analyzed using a mixed-effects model to explore overall effect of treatment relative to control as well as differences between treatments and the control group at each timepoint.

RNA extraction, sequencing, and de novo transcriptome assembly

RNA was extracted from samples of both species using the Qiagen RNEasy Plant Mini Kit and imaged on a 1.5% agarose gel for initial quality control. Additional quality control using BioAnalyzer (Agilent), library preparation and sequencing was performed at the University of Minnesota Genomics Center. Libraries were prepared using the TruSeq Stranded mRNA Library

Prep kit (Illumina), with Ampure-bead based size selection instead of gel-based. Pooled libraries were quality-checked by sequencing on the Illumina MiSeq platform and then sequenced using the S4 flow cell on the Illumina NovaSeq platform. Each species used an entire S4 flow cell. Sequence reads were paired end and 150 bp in length. RNA-Seq data quality was checked using FastQC v 0.11.7. Reads were trimmed of adapters and low quality sequences using Trimmomatic v 0.33 using the following parameters: ILLUMINACLIP:2:30:7 LEADING:10 TRAILING:10 SLIDINGWINDOW:4:20 MINLEN:35. Plastome sequences were removed by mapping reads to a publicly available conspecific representative plastome assembly (NCBI:NC_059015.1 for *B. maritima* and NCBI:ON598347.1 for *S. latifolia*) using Bowtie2 v 2.3.4.1 (Langmead and Salzberg 2012). The remaining reads were deduplicated using PRINSEQ-lite v 0.20.4 (Schmieder & Evans 2011), and transcriptomes were *de novo* assembled using TRINITY v 2.14.0 (Grabherr et al. 2011) with default parameters**.** For each species, four *de novo* transcriptomes were constructed, each from the aggregate of all sequencing libraries from across the time course for one particular treatment time course: control, high light, methyl jasmonate, and salt.

Pathway reconstructions

We carried out a literature review of functionally characterized genes from the pigment synthesis and related pathways in angiosperms, focusing on closely related Caryophyllales lineages where possible. We retrieved coding sequences (CDS) and protein sequences of these genes; protein sequences were used as baits to search the protein annotations of the reference genomes of *Beta* and *Silene* using BLASTp (Table S3.1) (Altschul et al., 1990). If no hits were found in the protein annotation, the reference genome CDS annotation and all four *de novo* assembled transcript datasets were searched using tBLASTn. Reference genomes of *Fagopyrum*

tataricum (CNCB:GWHBJBL00000000, He et al., 2023), *Simmondsia chinensis* (CNCB:GWHAASQ00000000, Sturtevant et al., 2020), *Mesembryanthemum crystallinum* (unpublished, J. Cushman), *Beta vulgaris ssp. vulgaris* (EL10.2, McGrath et al., 2023), and *Portulaca amilis* (v1.0, Gilman et al., 2022) were included in the search as outgroups to help evaluate orthologous and paralogous relationships among sequences. Top hits (high percent query sequence coverage and low e-value) were aligned with bait sequences using the auto option in MAFFT v 7.475 and maximum likelihood gene trees were inferred using default parameters and 1000 rapid bootstraps in IQ-TREE v1.6.12 (Katoh and Standley, 2013; Nguyen et al., 2015). We manually inspected alignments and gene trees to remove fragmented sequences and spuriously assembled sequences, and selected putative orthologs from alignments in an iterative process of tree inference and manual inspection. Additionally, MYB transcription factors were annotated from nucleotide reference and *de novo* assembled transcriptomes using MYB_annotator (Pucker 2022). MYB gene nomenclature follows the clades delimited in Stracke (2014). Clades with members that have an experimentally characterized regulatory role in flavonoid and phenylpropanoid metabolism were selected for further analysis.

Gene expression analyses

Transcript abundance was quantified using Salmon v 1.8.0 (Patro et al., 2015). Reads were mapped to reference transcriptomes from annotations of respective chromosome-level genome assemblies: *B. vulgaris* subsp. *maritima* genome version 1 (https://bvseq.boku.ac.at/Genome/Download/Bmar/; Dohm et al., 2014), and *Silene latifolia* genome from Yue et al. (2023; CNCB:GWHCBIJ00000000). Additionally, reference

transcriptomes for transcript quantification were supplemented with *de novo* transcripts of putative pathway genes discovered through pathway reconstruction described above.

DESeq2 v 1.38.3 (Love et al. 2014) was used to identify differentially expressed (DE) genes using a linear model accounting for the experimental design with a control time course (~treatment + time + treatment:time). Contrasts at each time point compared the expression of treatment groups to the control. DE genes were defined as having an absolute log-fold change > 2, and Benjamini-Hochberg-corrected p-value < 0.05.

Targeted metabolomics of Phe- and Tyr-derived compounds

Metabolites were extracted from samples after freeze-drying using a 2:1 HPLC-grade chloroform: Optima LCMS-grade methanol extraction solvent containing 0.5 ug/ ml isovitexin as an internal standard. 1 µL of each sample was eluted through a HSS T3 C18 reversed phase column (100 × 2.1 mm i.d., 1.8-µm particle size; Waters, Milford, USA) in a 26-minute gradient (Solvent A = 0.1% formic acid in water, solvent B= 0.1% formic acid in acetonitrile) with a flow rate of 0.4 mL/min and a temperature of 40° C. The binary linear gradient with following ratios of solvent B was used: 0-1 min, 1 %; 1-10 min, 1-10 %; 10-13 min, 10-25 %; 13-18 min, 25-99 %; 18-22 min, 99 %; 22-22 min, 99-1 %; 22-26 min, 1 %. Spectra were recorded in both positive and negative ion mode with a mass range of 100-1000 mz. The resolution was set at 140,000, and maximum scan time set to 200 ms. Sheath gas was set to a value of 45, while the auxiliary gas was set to 13. The transfer capillary temperature and heater temperature were both set to 350°C. The spray voltage was fixed at 3.75 kV. The tandem (MS/MS) spectra were recorded in negative and positive ion detection, covering a mass range from m/z 90 to 600 at a resolution of

70,000 and a maximum scan time set to 200 ms. Source gas settings were the same as for single MS mode.

Targeted tyrosine or phenylalanine-derived metabolites were identified in the spectra based on m/z and standards, and peak area was quantified from the extracted ion chromatograms. Peak area was normalized to mg fresh tissue weight. Isovitexin was used as an internal standard for correcting based on extraction efficiency. The targeted metabolites included L-Tyr, Phe, dopamine, L-DOPA, L(-)-epinephrine, rosmarinic acid, tyramine, L(-)-norepinephrine, homogentisate, N-methyl-4-tyramine, N-trans-feruloyltyramine, hordenine, 3-methoxytyramine HCl, salsolinol HBr, 4-hydroxyphenylpyruvate (4-HPP), 4-hydroxyphenyllactic acid (4-HPLA), salidroside, kaempferol, naringenin, quercetin, p-coumaric acid, sinapic acid, and caffeic acid. The standards for, N-trans-feruloyltyramine, salidroside, and salsolinol became unavailable after running the sea beet samples, but before running the *S. latifolia* samples, so they were unable to be quantified in *Silene*.

Reconstructed pathway gene-metabolite correlation

Gene-metabolite correlations were used to investigate the relationship between gene expression and downstream metabolite accumulation. Raw transcript counts were normalized using the 'scaledTPM' algorithm implemented in tximport v 1.26.1 (Soneson et al., 2015), which scales transcripts per million to library size and allows for cross-sample comparisons within the same experiment and species. For each treatment, Pearson correlation coefficients were computed to evaluate the relationship between reconstructed pathway gene expression and metabolite accumulation over the time course. In light of the co-option of *MYB21* from anthocyanin to betalain pathway regulation in table beet (Hatlestad et al., 2012), we also included putative MYB transcription factors that may be involved in betalain, flavonoid, or

phenylpropanoid metabolism in sea beet and *Silene*. Hierarchical clustering was used to group potential gene expression modules associated with metabolite biosynthesis.

RESULTS

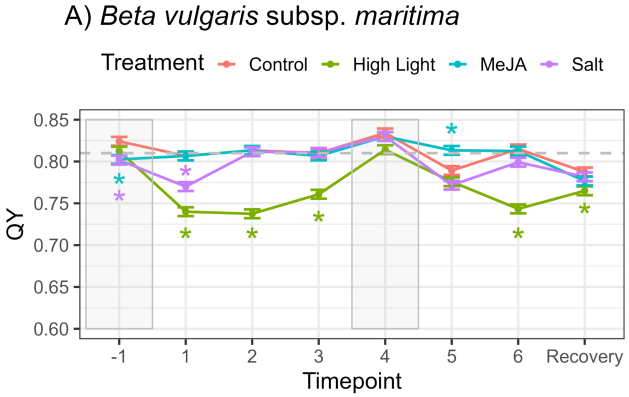
Physiology

In general, a quantum yield (QY) value at or above 0.83 is expected from a healthy, unstressed plant. A lower value indicates decreased photosynthetic capacity, usually due to inefficiencies at PSII brought on by environmental stress. In each species, the pre-experiment, pre-dawn QY values (Fv/Fm), which should be the plant's maximum, as all PSII reaction centers are open to receive electrons, were similar across the 24 plants for physiological measurements and ranged from 0.83-0.85 in *Silene* and 0.76-0.84 in sea beet, confirming the unstressed condition. High light caused the biggest drop in QY, with both methyl jasmonate and salt treatments having less impact on QY values (Fig 3.3). In both species, the high light-treated plants showed recovery in photosynthetic capacity at Time 4, which was the sole dark timepoint in the time course. A week after the experiment ended, the high light treatment plants had QY values that trended upwards again, although did not show a full recovery, likely due to changes in leaf structure and chemistry in response to high light conditions. One salt-treated plant in *Silene* also began dying during the recovery week, dragging that plant's QY down to 0.75. Generally, sea beet QY values were lower than in *Silene* for the control, salt, and MeJA time courses; however, Silene showed greater drops in QY values from high light treatment.

The PRI calculated from the hyperspectral scans also were the most impacted by high light treatments (Fig. 3.4)**.** The photochemical reflectance index reflects conformational changes in xanthophyll pigments and light use efficiency, and so is the index most comparable to QY measurements (Gamon et al., 1992). The NDVI is a measure of how green an area is, and can be

used as an estimate of plant health (Wang et al., 2016). No significant differences in treatment were apparent in the NDVI in either species (Fig 3.4**)**. The ARI was impacted by high light treatment in *Beta but* showed little change until the recovery time point in high light and salt plants in *S. latifolia* (Fig 3.4). ARI is an index designed for measuring red anthocyanin pigment, however, red betalains absorb at similar wavelengths (Gitelson et al., 2009). Pictures taken of each physiology group plant at each time point also show pigment visibly accumulating in the high light group of *Beta*, but not *S. latifolia,* with the exception of the dying salt plant. Neither MeJA nor salt treatments showed visible pigment accumulation (Fig 3.5).

Fig. 3.3-Quantum yield (QY) for each species over the time course by treatment. The x-axis is time point (Table 1). The -1 timepoint is the predawn pre-experiment measurement. In general, a QY value of 0.83 represents unstressed plants, and lower values indicate decreased photosynthetic capacity. High light had the greatest impact on both species. The dashed line represents the average of the control group at Time 0.



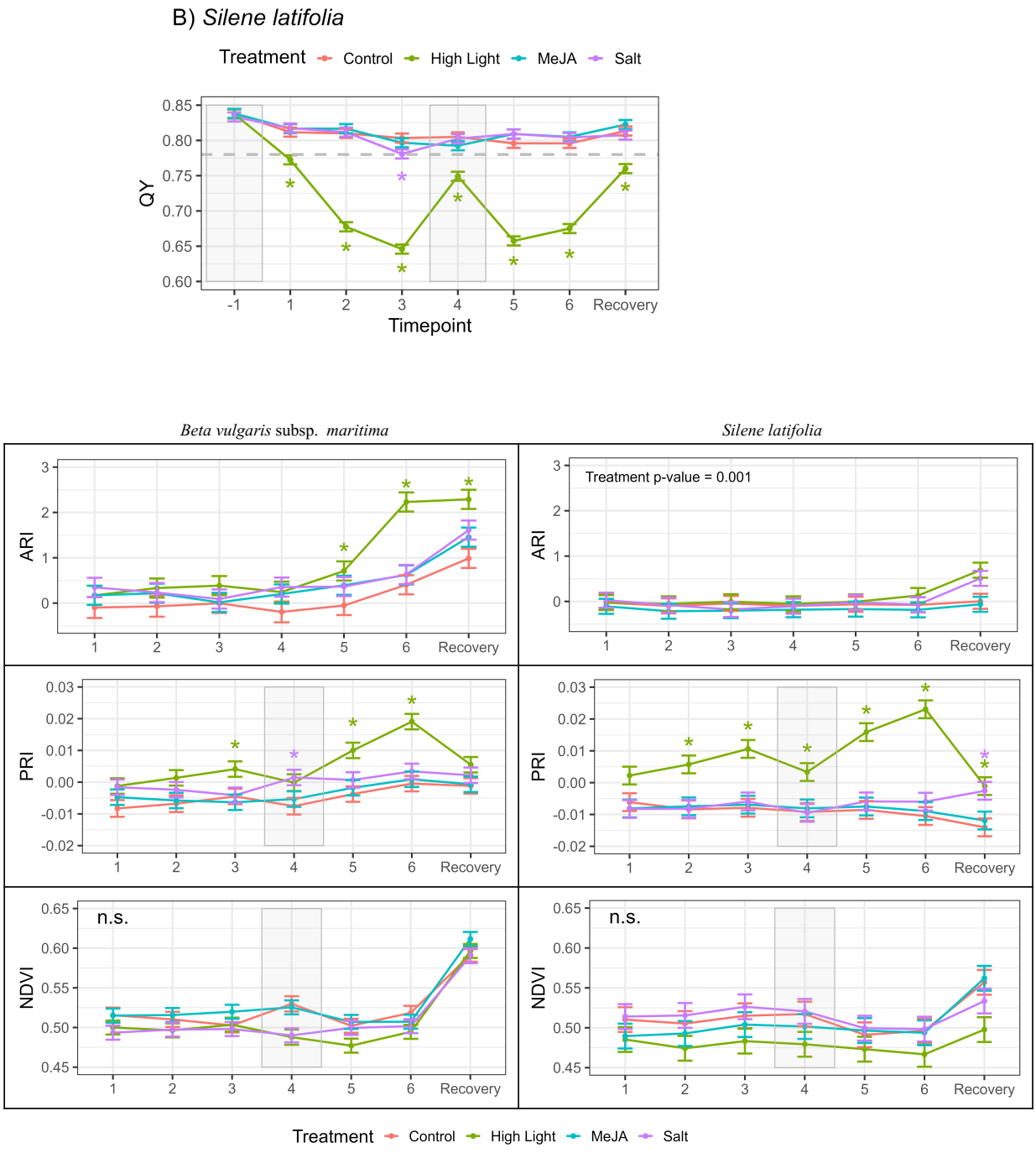


Fig 3.4- Hyperspectral indices (from top to bottom: ARI, PRI, and NDVI) by time point and species. The final time on the x-axis is recovery (one week post experiment). Colored asterisks above a time point indicate that treatment is significantly different from the control group at that timepoint. If treatments are not significant at specific timepoints, the p-value of overall treatment significance is given in the top left corner (n.s. = not significant).

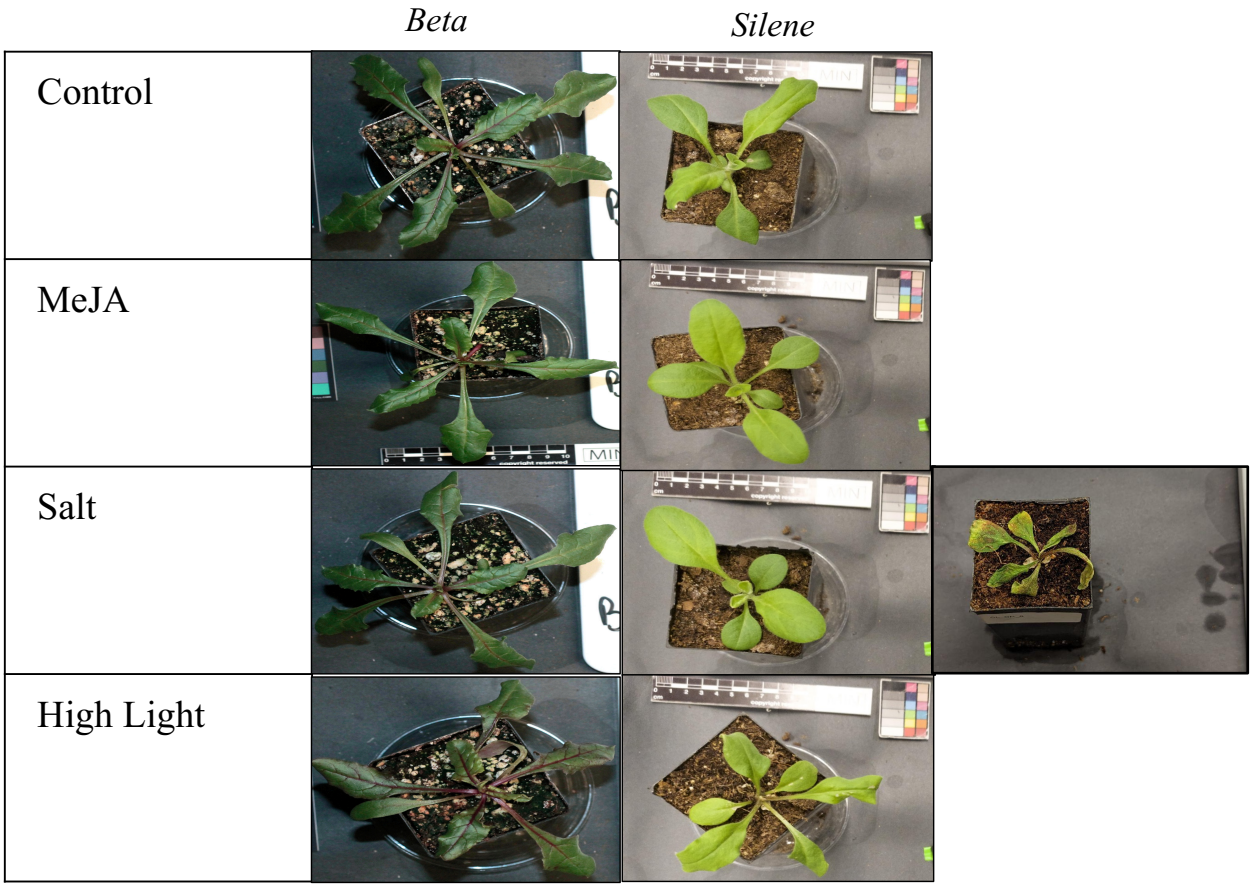


Fig. 3.5- Representative plants from each treatment group in both species at 48 hours (Time 6). The additional salt-treated *Silene* plant is from the Recovery timepoint, showing pigment accumulation as the plant dies.

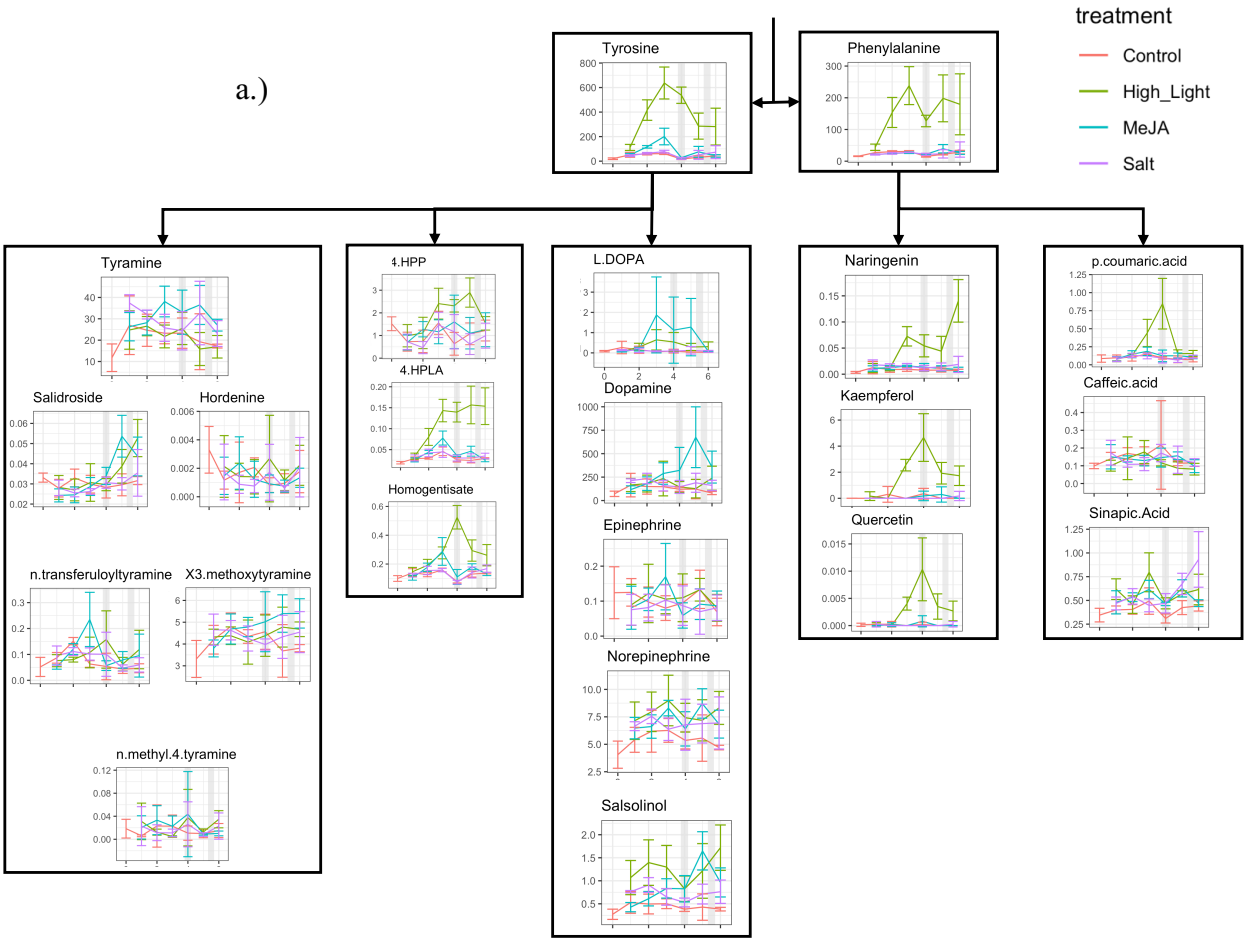
Metabolite profiling

Since it is impossible to quantify all possible end products, we chose metabolites that were situated at important branch points of the Tyr and Phe pathways. Of the quantified metabolites in *Beta*, 15 were tyrosine or tyrosine-derived, and 7 were phenylalanine or phenylalanine-derived. The concentration of both free Tyr and Phe strongly increased with high light treatment, with a 31.8 fold increase in Tyr and a 15.2 fold increase in Phe between the start of the experiment and their highest points at time point 3 (6 hours). The Phe-derived metabolites naringenin, p-coumaric acid, kaempferol and quercetin also responded to high light treatment, although they were all present at comparatively very low concentrations throughout. On the Tyr side, 4HPP, 4-HPLA, and homogentisate also were relatively low in concentration but positively responded to high light. The Tyr-derived metabolites L-DOPA (22.3 fold change) and dopamine (9.4 fold change) responded to MeJA treatments, while the Tyr metabolites salsolinol and salidroside responded positively to both high light and MeJA. Salt had no discernible effect on the measured metabolites in *Beta* (Fig 3.6)

In *Silene,* we were only able to quantify 12 out of 15 Tyr-derived metabolites due to the unavailability of standards. Of those, 4-HPP was not detected, and the rest did not have a noticeable response to treatment, with the exception of dopamine. On the phenylalanine side, Phe, naringenin and p-coumaric acid responded to high light treatment, although naringenin and p-coumaric acid were present in comparatively low abundance. The remainder of the Phe metabolites were not detected in any of the four reps at each timepoint, except for sinapic acid which is detected in a very low amount in the salt treatment plants at a single time point (Fig. 3.6).

In both species, the Tyr-derived tyramine and L-DOPA branches were highly active, with dopamine being especially abundant in both species. Tyramine also was even more abundant in *Silene* than in sea beet, peaking at an average of 38.1 nmol/gram in the sea beet's MeJA group at Time 3, and 1349.37 nmol/g in Silene's high light group at Time 3.

a.)



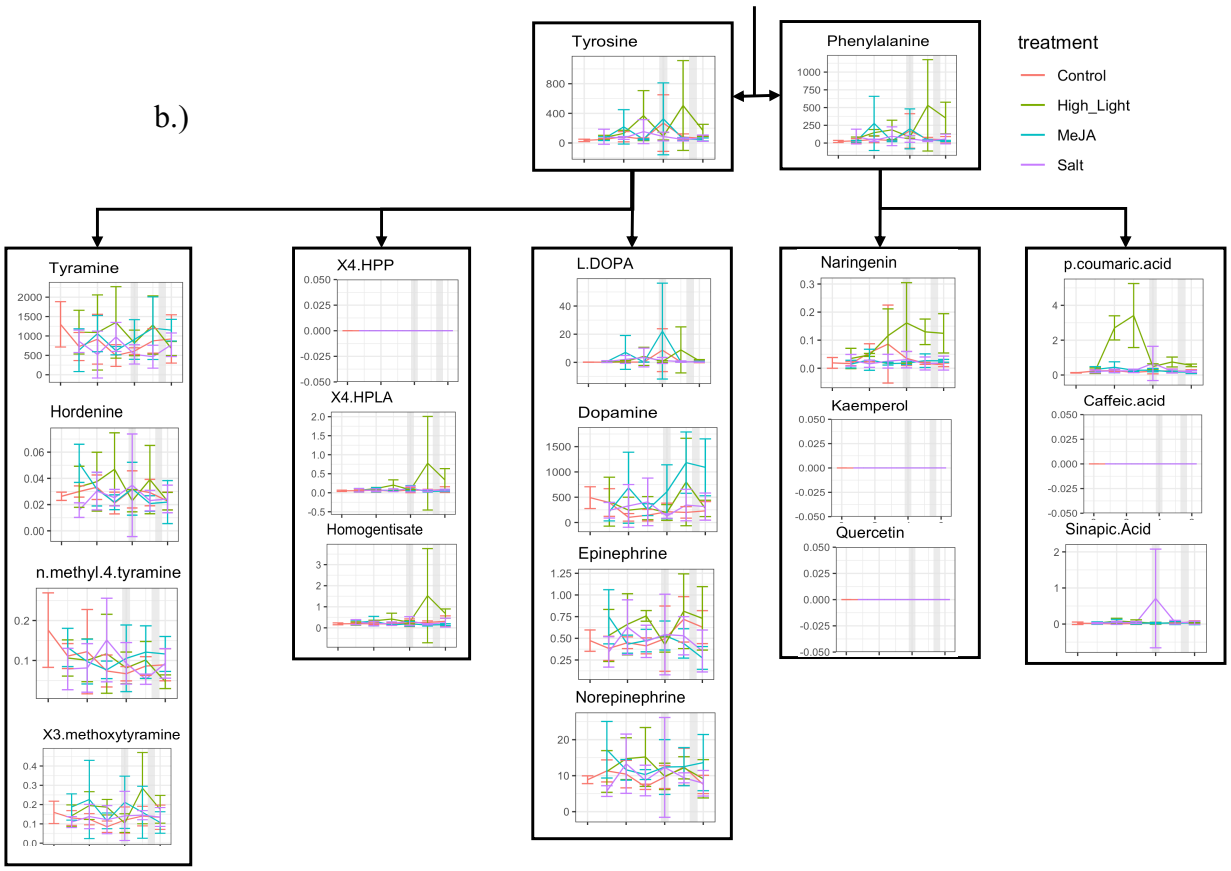


Fig. 3.6-Average metabolites by treatment over the course of the 48-hour experiment in a.) *Beta vulgaris* subsp. *maritima* and b.) *Silene latifolia*. Branches represent different pathways branching off the amino acid. In each graph, the x-axis represents timepoint. Time 4 (the midnight timepoint) and in between time 5 and 6 are shaded to represent nighttime. The y axis units are nmol/gram fresh tissue weight. Error bar indicates the standard deviation.

General patterns in differential gene expression

All three experimental treatment groups showed differentially expressed genes relative to the control at each time point in both species. Genes were generally upregulated in response to experimental treatments, except for in sea beet under the high light treatment during which a roughly proportional number of genes were found to be upregulated and downregulated. Of the three treatments, high light had the highest number of differentially expressed genes. Within the high light treatments, responses peaked with the most DEGs at Time 5 in sea beet (24 hours, 878 genes) and Time 3 in *Silene* (6 hours, 495 genes). Methyl jasmonate treatment had the most DEGs in both species at Time 1 (1 hour) with 268 genes in sea beet and 339 in *Silene* (Fig. 3.7). There were fewer differentially expressed genes (DEGs) in salt compared to the other treatments, with a slight response that peaked at Time 6 (48 hours) for *Beta* and Time 3 (6 hours) for *Silene*; there were 17 DEGs in beet and 40 in *Silene* at their peak differential expression.

Very few genes from the reconstructed pathways were found to be differentially expressed in either species, and treatments differed in the extent of pathway genes found to be differentially expressed. In general, all reconstructed pathway DEGs were upregulated except for a putative THT downregulated at Time 6 in the *Silene* salt treatment, and a putative phenylpropanoid repressor MYB14 at Time 1 in the sea beet high light treatment (Table S3.2). Reconstructed pathway DEGs follow the same trends as the total DEGs described above except

for in the methyl jasmonate treatment in *Silene*, during which DEGs from the pathway reconstructions peaked at Time 2 (3 hours, 6 genes) (Fig. 3.7).

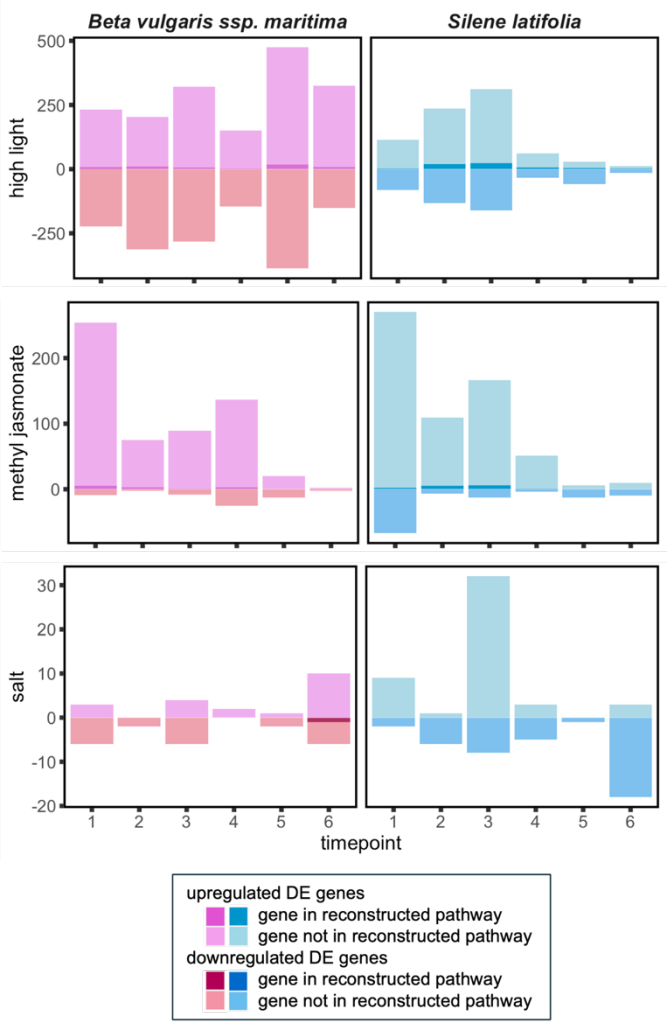


Figure 3.7. Total number of genes differentially expressed at each timepoint for each treatment compared to the control time course. Upregulated gene counts are shown above the x-axis; downregulated gene counts are shown below. Dark-colored bars overlaid on light-colored bars indicate the total number of genes annotated in reconstructed pathways found to be differentially expressed.

Shikimate pathway gene expression

Expression of genes in the shikimate pathway, which produces arogenate, the precursor to Tyr and Phe, was also plotted (Fig. 3.8). Overall, the shikimate pathway genes were up-regulated in response to MeJA treatment, particularly in sea beet. High light treatment also induced

expression of most genes, although typically to a lesser degree than MeJA. In a few cases (DAHPS in *Silene,* DHQ/SDH in sea beet, CM in both) one copy of the gene was MeJA responsive, and the other high light responsive. Gene expression in the salt treatment did not differ significantly from the control.

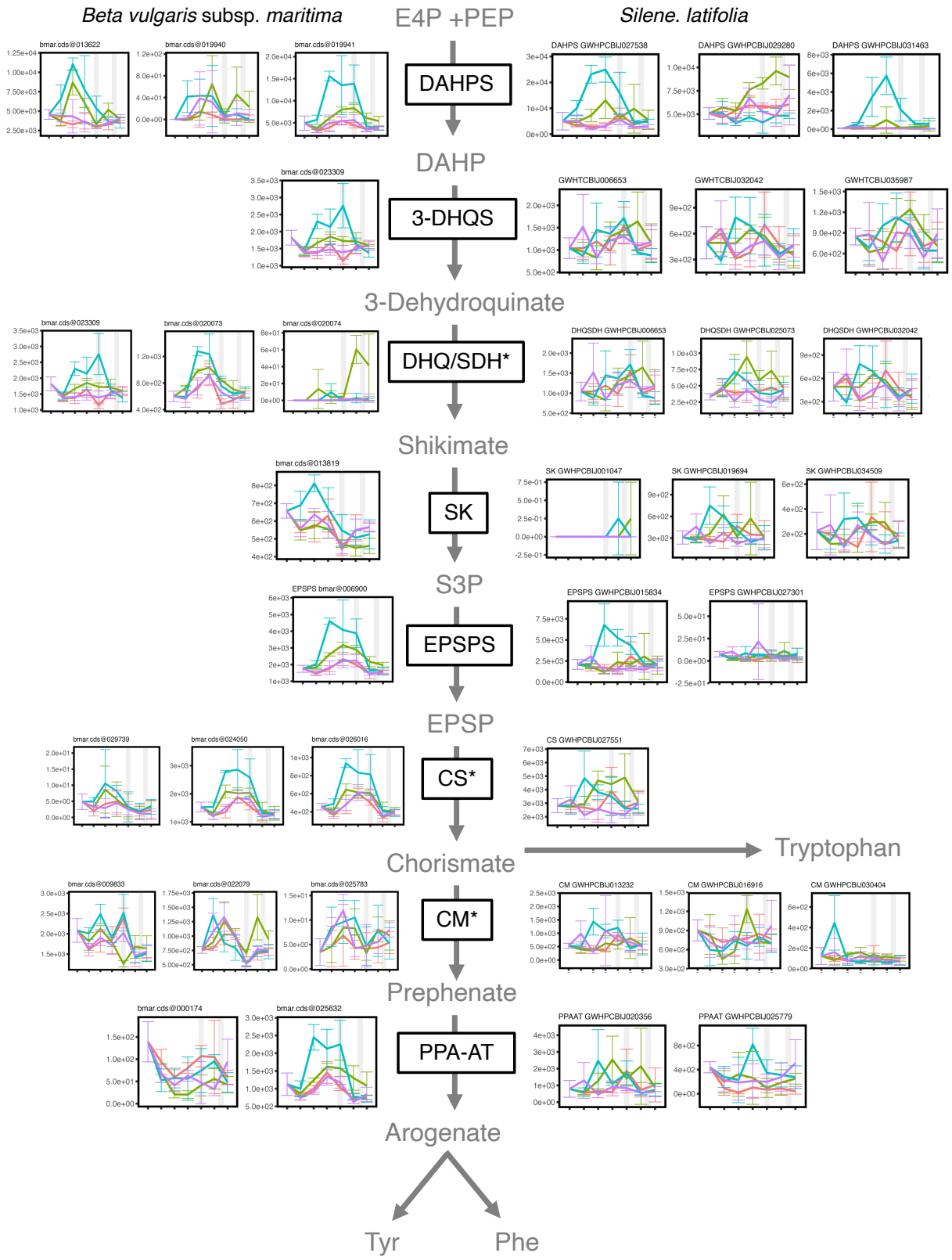

Beta vulgaris subsp. maritima
Silene. latifolia
E4P +PEP
DAHPS
DAHP
3-DHQS
3-Dehydroquinate
DHQ/SDH*
Shikimate
SK
S3P
EPSPS
EPSP
CS*
Chorismate
Tryptophan
CM*
Prephenate
PPA-AT
Arogenate
Tyr
Phe

Fig 3.8-Gene expression in the shikimate pathway of both sea beet (left) and *Silene* (right). Metabolites in the central pathway are in gray font, and the gene names are in black font in black boxes. Asterisks next to a gene name indicate additional copies were found and can be seen in Fig. S3.1.

Pigment Pathway and Branching Pathway Genes

Many of the genes involved in betalain synthesis and modifications were recovered in *Silene*, despite not being a betalain producer (Fig. 3.9). Notably missing are *CYP76AD1α* and *DODAα*, which are both deterministic in betalain production. *CYP76AD1α* is one of three clades of *CYP76AD1* paralogs. Notably, members of the other two clades, *β* and *γ,* were also not found in *Silene.* However, paralogs of *DODAα* genes were recovered. The paralogs belong to the *DODAβ* clade, which has no demonstrated betalain synthesis activity (Fig. 3.10; Brockington et al., 2015; Sheehan et al., 2020). Also not recovered, consistent with previous analyses, is *ADHa* (Lopez-Nieves et al. 2018).

The high light treatment groups show similar patterns of gene expression shared by both species **(**Fig 3.9**)**. All the early anthocyanin genes have at least one copy that responds positively to high light. These light responsive-early genes also have nearly identical expression patterns, although the actual magnitude of expression varies. Late genes, which are deterministic for anthocyanin, with the exception of *UF3GT* in both species and *AT* in *Silene*, have little to no expression in both species. In the betalain pathway, one copy each of *CYP76AD1α* and *DODAα,* which were only recovered in sea beet, also respond positively to high light, with a nearly identical expression pattern to each other. Both of these high light-responsive copies are orthologs of previously functionally characterized betalain-producing genes in *Beta vulgaris* (Hatlestad et al., 2012). Additionally, the *MYB21* regulatory gene known to have been co-opted from anthocyanin synthesis in *Beta* was found in both species and has a nearly identical

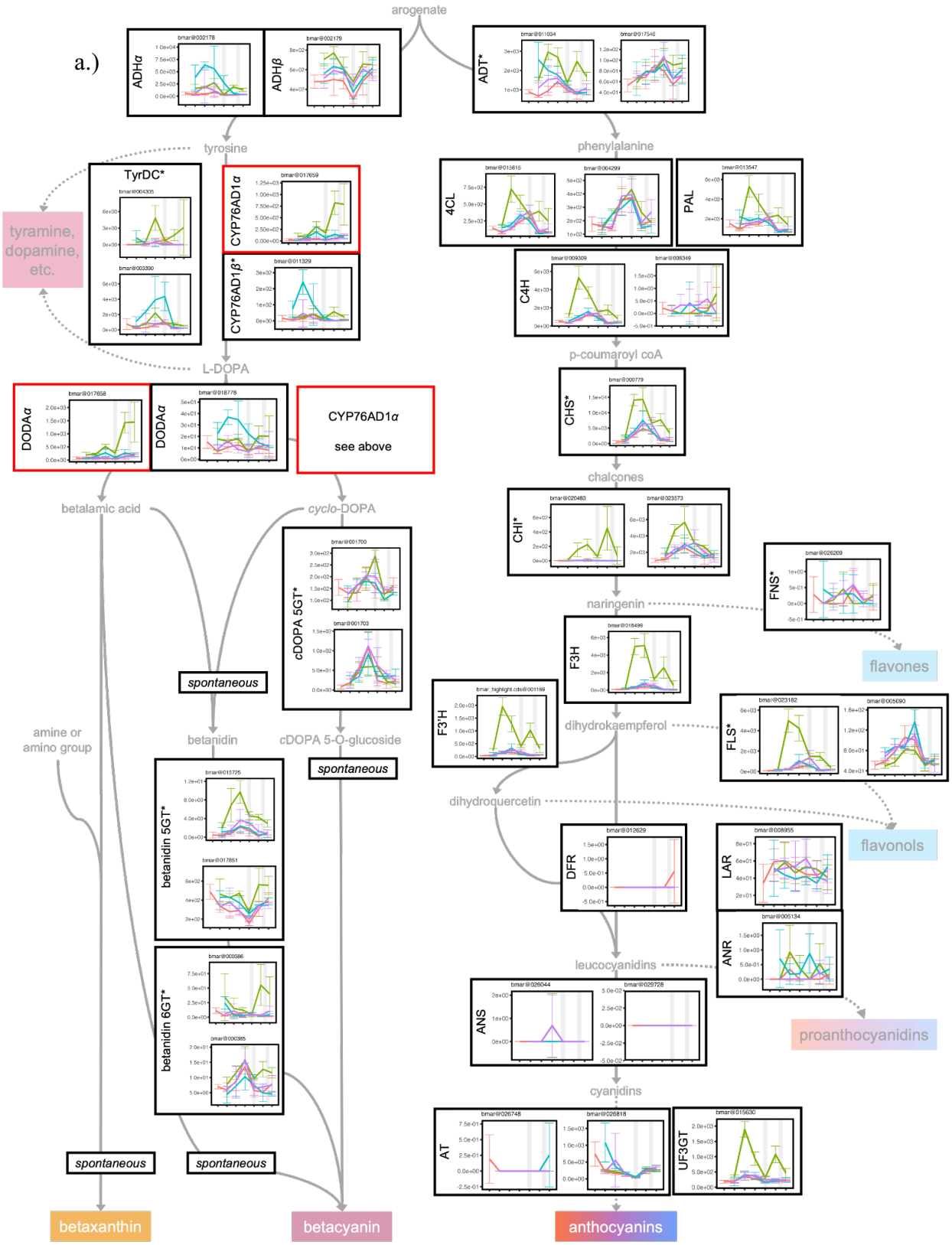
expression pattern to the light-responsive *CYP76AD1α* and *DODAα* in *Beta* (Fig 3.10). *CYP76AD1β,* a paralog of *CYP76AD1α* that converts Tyr to L-DOPA and was only recovered in sea beet, had an early response to MeJA and a later response to high light. Other tyrosine-related genes do not respond to high light consistently between species.

Several of the betalain pathway genes respond most strongly to MeJA treatment, including all three members of *CYP76AD1β* and two of the six non-betalain-producing members of the *DODAα* clade in sea beet. Notably, *ADHα* in sea beet and *ADHβ* in *Silene* share very similar expression patterns, which respond positively to MeJA and minimally to other treatments. In contrast, expression of sea beet's copy of *ADHβ* mostly varied between day and nighttime, with slight increase under high light

Other genes that branch off the pigment pathways also respond to either high light or MeJA. Both *Silene* and sea beet *TyrDC* respond to MeJA, although the gene also has a light responsive copy in sea beet. *THT,* a gene family which codes for the enzymes responsible for converting tyramine into several downstream metabolites, has both light responsive and MeJA responsive copies in sea beet (Fig S3.1). On the phenylalanine side, *FNS*, which synthesizes flavones, appears to be lost in *Silene*, and is unresponsive to treatment in sea beet. *FLS*, which is responsible for flavonol synthesis, is present in both species but only has a treatment-responsive copy in sea beet. Both species have high light-responsive copies of *HCT*, which is one of the first steps towards lignin synthesis.

Overall, Phe pathway genes, especially early ones, respond primarily to high light. Tyr pathway genes respond to both high light and MeJA treatment. Similarly to the patterns for metabolites, salt treatment has little impact on the expression of pathway genes.

Despite the seeming lack of change in pathway genes in the salt treatment, pathway genes and identified *MYB* genes do form correlation modules with the metabolites distinct from the control group in both species, as do the high light and MeJA groups (Fig 3.11).


a.)
arogenate
ADHα
ADHβ
ADT*
tyrosine
phenylalanine
TyrDC*
CYP76AD1α
CYP76AD1β*
4CL
PAL
C4H
tyramine, dopamine, etc.
L-DOPA
p-coumaroyl coA
DODAα
DODAα
CYP76AD1α
see above
CHS*
chalcones
betalamic acid
cyclo-DOPA
CHI*
cDOPA 5GT*
naringenin
FNS*
flavones
F3H
F3'H
spontaneous
dihydrokaempferol
FLS*
amine or amino group
betanidin
cDOPA 5-O-glucoside
spontaneous
dihydroquercetin
flavonols
DFR
LAR
betanidin 5GT*
leucocyanidins
ANR
betanidin 6GT*
ANS
proanthocyanidins
cyanidins
AT
UF3GT
spontaneous
spontaneous
betaxanthin
betacyanin
anthocyanins

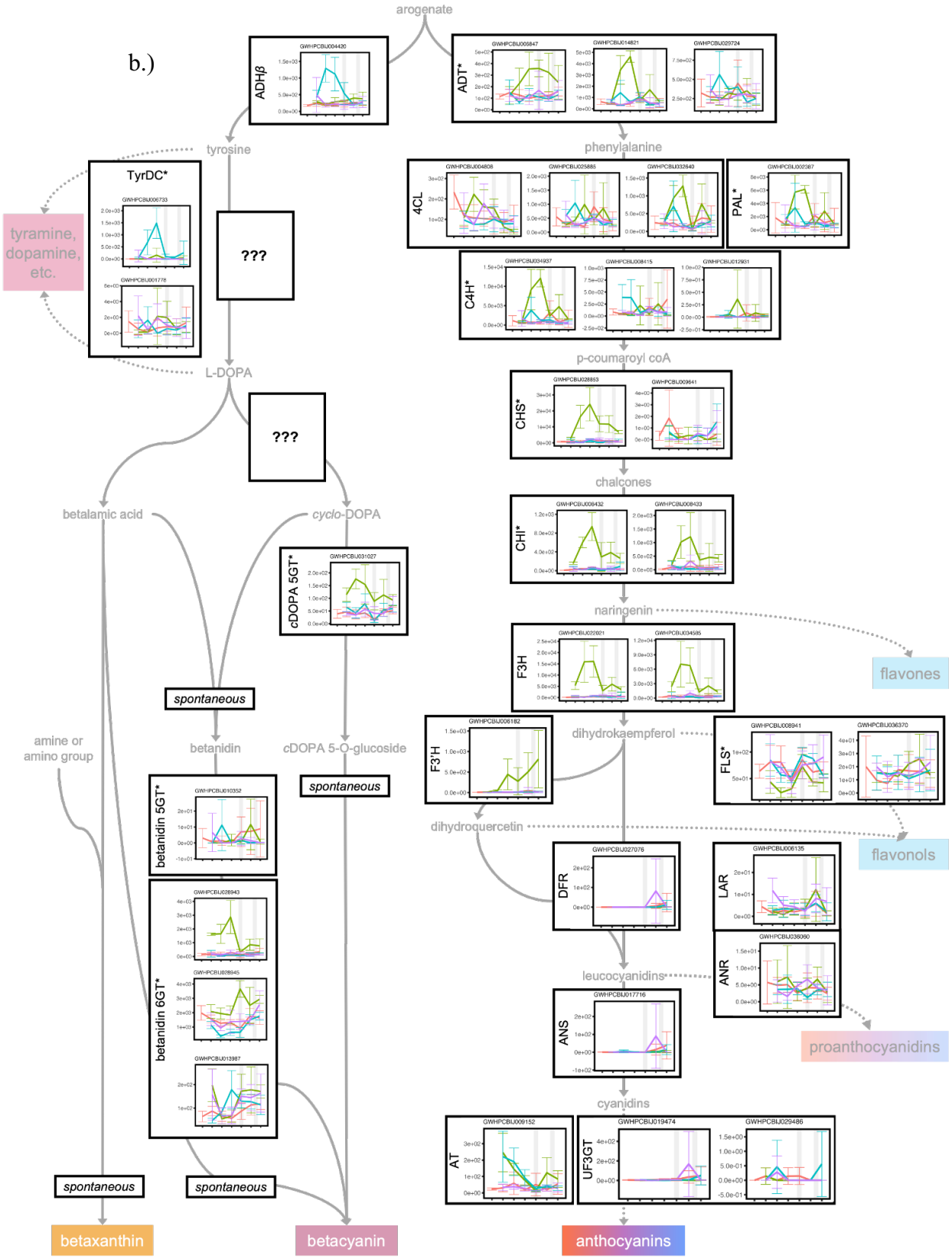

b.)
arogenate
ADHβ
ADT*
tyrosine
phenylalanine
TyrDC*
tyramine, dopamine, etc.
???
4CL
PAL*
C4H*
p-coumaroyl coA
L-DOPA
???
CHS*
chalcones
betalamic acid
cyclo-DOPA
cDOPA 5GT*
CHI*
naringenin
F3H
flavones
spontaneous
dihydrokaempferol
F3'H
FLS*
amine or amino group
betanidin
cDOPA 5-O-glucoside
spontaneous
dihydroquercetin
flavonols
betanidin 5GT*
DFR
LAR
betanidin 6GT*
leucocyanidins
ANR
ANS
proanthocyanidins
cyanidins
AT
UF3GT
spontaneous
spontaneous
betaxanthin
betacyanin
anthocyanins

Fig 3.9- Anthocyanin, betalain, and selected notable branch pathway gene expression of a.) *Beta vulgaris* subsp. *maritima* and b.) *Silene latifolia* over the experimental time course. Asterisks by the gene name indicate there are copies of that gene recovered but not shown in the figure; selected genes show representative patterns of gene expression and response to experimental treatments. Homologs to functionally validated betalain pathway genes are highlighted in red. Expression of all recovered pathway gene copies can be found in Fig S3.1.

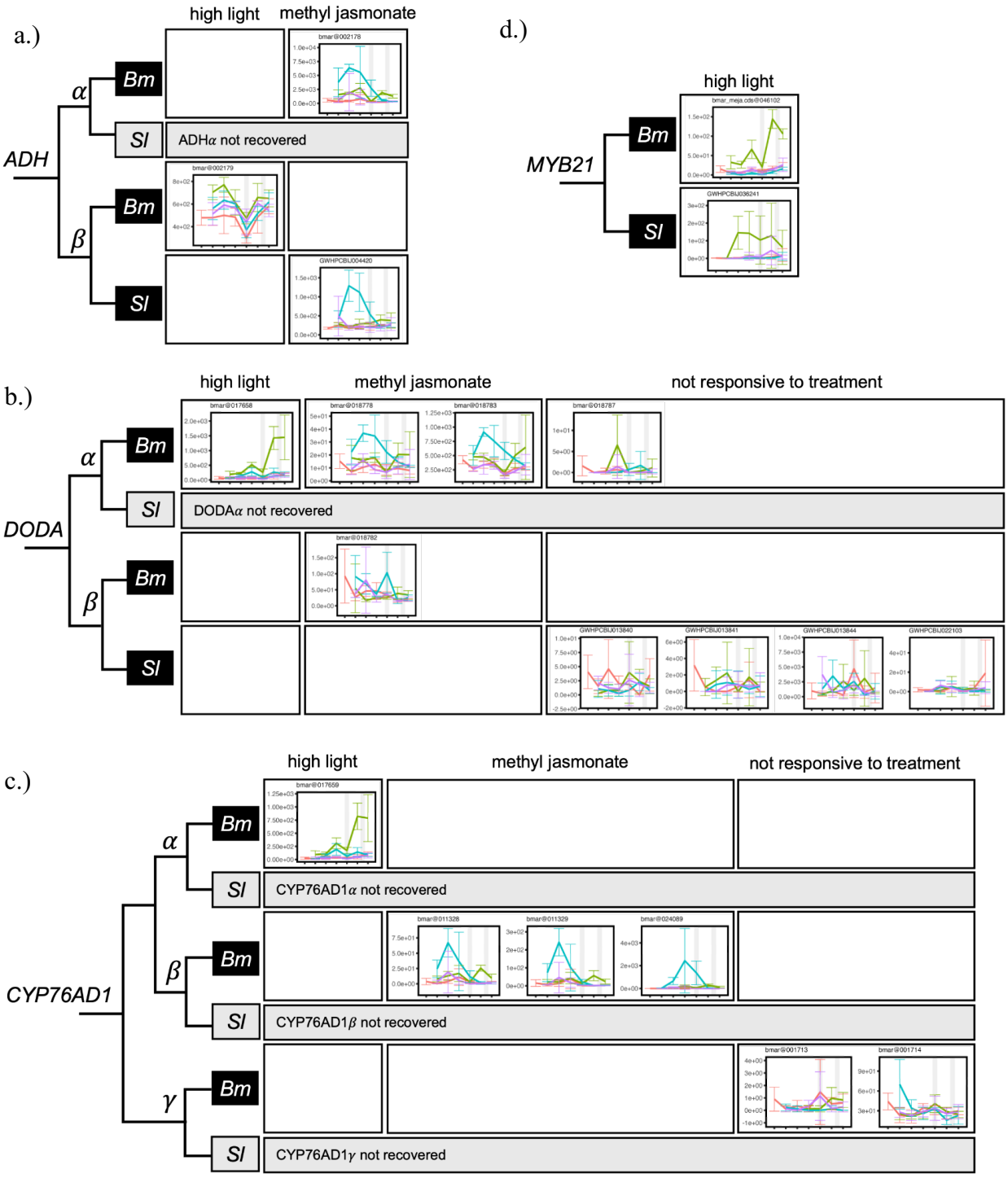
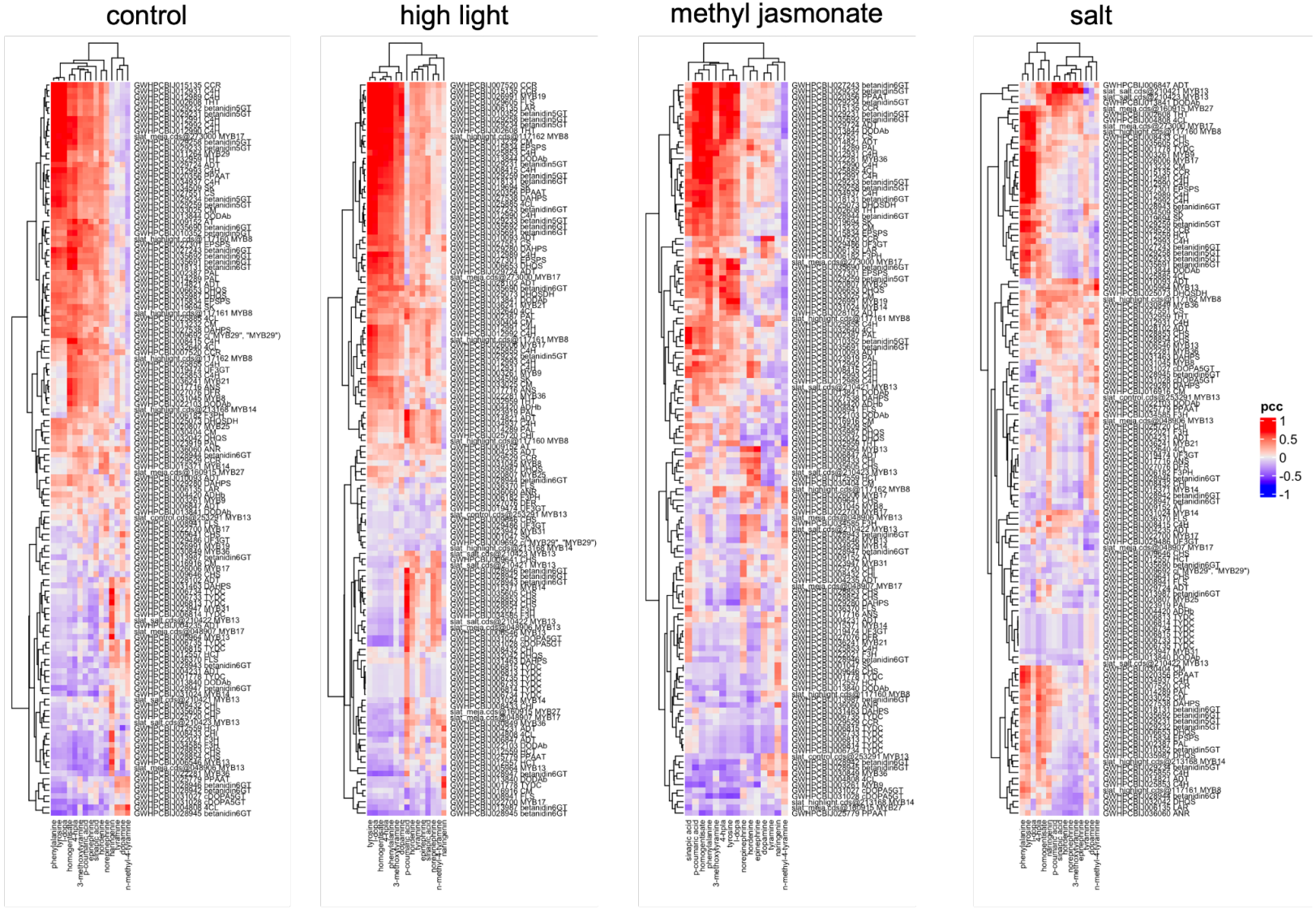

a.)
high light
methyl jasmonate
ADH
α
β
Bm
Sl
ADHα not recovered
d.)
high light
MYB21
Bm
Sl
b.)
high light
methyl jasmonate
not responsive to treatment
DODA
α
β
Bm
Sl
DODAα not recovered
c.)
high light
methyl jasmonate
not responsive to treatment
CYP76AD1
α
β
γ
Bm
Sl
CYP76AD1α not recovered
CYP76AD1β not recovered
CYP76AD1γ not recovered

Figure 3.10. Comparative responses of all paralogs of focal betalain pathway genes (A) *ADH*, (B) *DODA*, (C) *CYP76AD1* and (D) *MYB21* in *Beta vulgaris* subsp. *maritima* (*Bm*) and *Silene latifolia* (*Sl*). Gene copies are categorized by their response to treatment. The absence of *S. latifolia* copies for *ADHa, DODAa,* and *CYP76AD1* is noted in dark boxes.

Silene latifolia



Beta vulgaris subsp. *maritima*

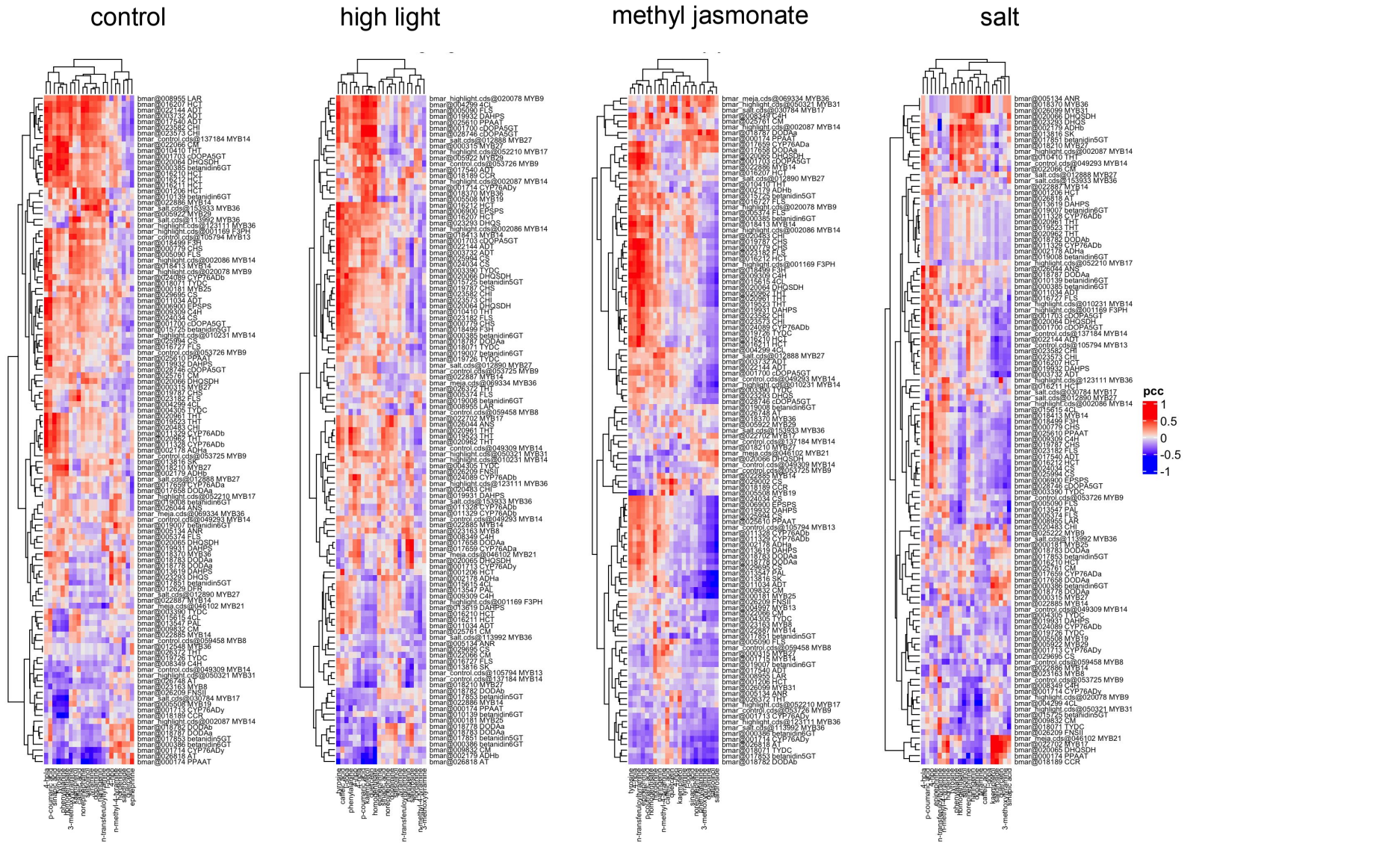


Fig 3.11- Gene-metabolite correlations of reconstructed pathway genes and annotated MYB genes with targeted metabolites for (A) *Silene latifolia* and (B) *Beta vulgaris* subsp. *maritima*. Color indicates Pearson correlation coefficient. Cluster dendrograms along the rows and columns show potential modules of correlated gene expression and metabolite accumulation.

DISCUSSION

Hyperspectral methods provide rich information to measure stress

The photochemical reflectance index (PRI) calculated from the hyperspectral scans performed comparably to the QY measurements in assessing stress through photosynthetic capacity. High light treatment is known to damage photosystem II, so its effect on QY and PRI is expected (Murata et al., 2012). Neither showed much of a response in the MeJA or salt treatments. The exceptions are the recovery timepoint in salt, where one of the measured *Silene* plants was dying, and one mid-time course salt point significantly differing in QY in *Silene* and PRI in *Beta*. It is possible that neither treatment induced a response in the plants that significantly affected photosynthesis, although MeJA has been shown to affect photosynthetic efficiency in other species (Kurowska et al. 2020). On the other hand, NDVI did not show any significant differences between treatments, indicating that the treatments did not affect chlorophyll pigmentation. The anthocyanin reflectance index (ARI) detected pigment accumulation in sea beet, despite the pigment being betalain rather than anthocyanin. It also confirmed visual observations that *Silene* was not significantly accumulating pigment at any timepoint until after treatments ended, although overall treatments were significantly different from controls. The slight uptick in ARI in the high light and salt points at Time 6 and Recovery may be contributing to this, and suggest anthocyanin was accumulating in these plants. This index was designed to target red anthocyanins, which reflect at similar wavelengths to the red sub-class of betalains, betacyanins (Gitelson et al., 2009). While the ARI works well in this

instance for estimating the trend of betalain accumulation in sea beet, which is primarily betacyanin accumulating, it would be less applicable in a species that primarily accumulates the yellow betaxanthins.

Overall, spectral data provided more information than QY on stress responses. Multiple different indices providing different information were calculated from the same scan, and the spectral data was better able to capture the changes in salt treatment during recovery. However, QY measurements connect more directly with plant physiology, and require less processing to analyze the data than spectra do. Additionally, the spectral set up is expensive, requires training, and requires significant data storage space, as the files for this experiment come to 100 GB.

Salt treatment had a relatively weak response in our focal pathways

In the metabolomics data, physiological measurements and pathway gene expression, virtually no effect of salt treatment was observed, although spectral indices showed a small salt response in later timepoints in *Silene*. Similarly, in the global differential expression analysis, far fewer genes were differentially expressed in salt treatments relative to the controls. This relatively weak response is unlikely to be due to using too low of a salt concentration. Before the experiment we tested a range of salt concentrations in both species and chose the highest salt concentration that didn't induce a loss of turgor pressure over 48 hours.

Part of the weak salt response may be due to the plant species chosen. Sea beet is a halophyte that is well-adapted to salty conditions and may not need to induce genes in response to salt, instead having genes contributing to salt tolerance nearly always "turned on". *Silene latifolia* is not a halophyte, but does prefer disturbed areas and is a common weed in areas of human habitation. Its preference for these disturbed areas may confer more salt stress tolerance than a typical non-halophyte. We do see that *Silene* has a stronger response than the halophytic

Beta in the number of differentially expressed genes, particularly at Times 3 and 6. The early induction of differential expression at Time 3 is consistent with response patterns in the MeJA and high light treatments.

Alternatively, salt stress may be a slower-acting stress than MeJA and high light, and our experimental time frame is too short to capture its effect. This may be evident in the uptick in differentially expressed genes at the final time point at Time 6 (48 hours). Following the salt treatment, one *Silene* plant also died after the recovery measurements. Salt was administered by soaking the soil for an hour at the beginning of the time course, with no more watering during the time course, so the salt concentration in the soil increased over time, contributing to its later response. Salt plants also correlated to metabolites in different patterns than the control plants or the other two treatments (Fig. 3.11).

The relatively weak response to salt treatment by pathway genes and metabolites, despite global differential gene expression occurring relative to the control, suggests that salt treatment may affect pathways other than the ones focused on here. BLAST searches of salt DEGs (Table S3.3) against *Arabidopsis thaliana* suggest that the early stages of salt stress response in both *Silene* and beet are mainly concerned with slowing growth and dealing with imbalances in cellular turgor pressure, rather than inducing some of the potential photoprotective properties of our focal pathways (Jain and Gould 2015). Beet also upregulates genes associated with pathogen resistance. These findings are consistent with studies on similar experimental timescales in closely related Amaranthaceae halophytes, *Chenopodium quinoa* Willd. (Shi and Gu 2020) and *Salicornia europaea* L. (Ma et al., 2013), in which very few pigment, phenylpropanoid, and oxidative stress-related genes were found to be differentially expressed relative to many cell growth and ion stress-related genes.

Expression in the shikimate pathway matches canonical expectations

The shikimate pathway, while leading into specialized metabolism, is itself a part of primary metabolism, as it synthesizes the three aromatic amino acids: phenylalanine, tyrosine, and tryptophan. In *A. thaliana*, the first step in the pathway begins with the synthesis of 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) by DAHP synthase (DAHPS). One of three DAHPS isoforms is feedback inhibited by excess tyrosine, while the other two are associated with anthocyanin accumulation under high light conditions (Yokoyama et al., 2021). Here we recovered three DAHPSs in each species. Two in each species are up-regulated in response to MeJA. The third is most up-regulated by high light, although MeJA still has a positive response in *Beta*. Similarly, it has been shown in *Arabidopsis* that this gene has copies that are up-regulated by biotic stress and copies that are not (Keith et al. 1991). Additionally, in tomatoes, it has been shown that the *EPSPS* and *CS* genes also positively respond to biotic factors (Görlach et al., 1995). Overall, the shikimate pathway in *Silene* is not unusual relative to other species, implying that the unusual downstream gene expression and metabolite accumulation we see in the species is restricted to the Tyr and Phe pathways and not a more widespread shift in aromatic amino acid metabolism.

The accumulation of Tyr and Phe show similar patterns to each other in both species, and also cluster together in nearly all gene-metabolite correlation models, with the exception of high light *Beta* (Fig. 3.6; 3.11). Some of the accumulation of the free amino acids can probably be attributed to protein degradation, particularly under high light conditions. However, they also tend to correlate highly with shikimate pathway genes, suggesting that the similar patterns of expression are partially attributable to their upstream pathway.

High light and methyl jasmonate impact on pigment genes and metabolites

High light and MeJA treatments had different effects on gene expression between the anthocyanin and betalain pathways. MeJA impacted betalain/wider tyrosine gene expression more, while high light had a greater impact an anthocyanin/phenylalanine genes, although a few betalain genes were also upregulated by high light. The Tyr genes that responded to high light in sea beet, *CYP76AD1α* and *DODAα,* had a nearly identical pattern of expression suggesting that they were co-regulated. Given that the rest of the betalain genes either respond to MeJA or none of the treatments and the early anthocyanin pathway is high light responsive, it is possible the regulator responsible for *CYP76AD1α* and *DODAα* in sea beet was co-opted from Phe metabolism, explaining their responsiveness to high light. Notably, *MYB21*, the regulatory gene co-opted from anthocyanin synthesis shares a very similar expression pattern to these two genes and may play a role in regulating them. Similarly, *cDOPA5GT* in *Silene* is responsive to high light with a similar expression pattern to the early anthocyanin genes. While it's unclear what this gene is doing in *Silene*, it may be co-regulated with the early anthocyanin genes.

The high light responsive early anthocyanin synthesis genes also had a nearly identical pattern to each other and are likely to share a regulator. The Phe-derived metabolites, when detected and treatment responsive, also respond to high light. Although they don’t share quite the same pattern as gene expression, kaempferol and quercetin in *Beta* especially appear to have a time-delayed similar pattern to the early anthocyanin genes in response to high light.

The pattern of Tyr genes and metabolites more associated with MeJA and Phe genes and metabolites more associated with high light suggests that sea beet’s Tyr metabolism is not functionally homologous to Phe metabolism in *Silene* as predicted. Instead, at least in sea beet, betalain and wider Tyr metabolism seems to have been shaped more by biotic stresses like

herbivory, while anthocyanin and wider Phe metabolism in both species have been influenced by abiotic forces like light stress. Salt stress, on the other hand, preferentially affects pathways not investigated in this study.

Expression of ADHβ in Silene was similar to the deregulated ADHα in sea beet

One of the most surprising results is the pattern of gene expression in *ADH*. We expected that, given the deregulation of *ADHα* and its association with betalain synthesis, we expected it to respond to treatment in sea beet, which it did. On the other hand, we expected the canonical, feedback-regulated, non-pigment-associated *ADHβ* to be less responsive and consecutively expressed at a low level in both species. While this was the case in sea beet, in *Silene,* the only copy of *ADH*, the canonical *ADHβ,* had an expression pattern more like the *ADHα* copy in sea beet, being strongly induced by MeJA treatment, especially at the 3-hr and 6-hr timepoints. Phylogenetic analysis confirmed that *Silene latifolia* copy of *ADH* is the *β* copy, forming a clade with other *ADHβ* copies in Caryophyllales (Fig 3.10). Ancestors of Caryophyllaceae did have a copy of *ADHα*, before it was lost in the basal grade of Caryophyllaceae (Lopez-Nieves et al., 2018). Given the abundance of Tyr-derived metabolites in *Silene*, as well as the expression of other Tyr metabolism genes, it is possible that *ADHβ* has taken on some of the ancestral *ADHα* activity to compensate for its loss and feed to downstream pathways. Alternatively, *ADHα* and *ADHβ* may be regulated similarly across Caryophyllales and *ADHβ* in sea beet has undergone a shift of function. Wider species sampling is needed to tease apart these alternative scenarios.

Losses of multiple betalain synthesis genes in Silene latifolia

Despite not being a betalain producer, we recovered several of the betalain pathway genes in *Silene*, including *cDOPA5GT*, *betanidin5GT*, and *betanidin6GT.* Notably missing were

CYP76AD1 and *DODAa*, which code for enzymes that catalyze two of the deterministic steps in betalain synthesis. Both genes are part of larger gene families, and have undergone Caryophyllales-specific duplications early in core Caryophyllales that led to acquiring betalain synthesis activity.

In *DODA*, the early core Caryophyllales gene duplication resulted in the *DODAa* and *DODAβ* clades (Brockington et al., 2015; Sheehan et al., 2020). Further duplication in the *DODAa* clade led to seven *DODAa* copies, one of which acquired high levels of L-DOPA 4,5 dioxygenase activity that converts L-DOPA to betalamic acid (Sheehan et al., 2020). Although not all *DODAa* copies have high L-DOPA 4,5 dioxygenase activity, no copies in the *DODAβ* clade seem to have such activity. While *Silene* has several copies of *DODAβ,* it does not have any *DODAa.*

CYP76AD1 is part of the large cytochrome P450 family. Within *CYP76AD1*, gene duplication led to three clades: *CYP76AD1α*, *CYP76AD1β,* and *CYP76AD1γ* (Brockington et al., 2015). Several members of the *α* clade across Caryophyllales species have demonstrated L-DOPA oxidase activity, catalyzing the conversion of L-DOPA into cyclo-DOPA (Brockington et al., 2015; Hatlestad et al. 2012; Sunnadeniya et al., 2016). Members of the *α* clade are also capable of catalyzing the previous step, hydroxylating tyrosine into L-DOPA. However, the *β* clade has only demonstrated activity in the previous step (tyrosine to L-DOPA). The *γ* clade has no known betalain synthesis activity. While sea beet has *CYP76AD1α*, *CYP76AD1β,* and *CYP76AD1γ*, none were recovered in *Silene latifolia.* In fact, despite searching both *de novo* assembled transcriptomes and the reference genome using tBLASTn and BLASTp, we could not find any member of *CYP76AD1* in *Silene,* suggesting there was a complete loss of *CYP76AD1* in Silene. This is especially surprising considering the metabolites L-DOPA and dopamine were

detected in the species. In fact, dopamine in the MeJA treated plants is in comparatively high abundance to other measured metabolites in the species. However, the presence of these metabolites would normally require the tyrosine hydroxylation step *CYP76AD1α/β* provides. While *CYP76AD1α* has yet to be found in an anthocyanin lineage, *CYP76AD1β* has been in other members of the Caryophyllaceae family, meaning this loss may be unique to *Silene* (Brockington et al., 2015). If there truly was a loss of *CYP76AD1* in *Silene,* another gene must be responsible for this step. The cytochrome P450 gene family is large and promiscuous, so it is possible a different *CYP* has evolved this function in *Silene* (Hansen et al., 2021). Alternatively, before the discovery of *CYP76AD1'*s role in betalain synthesis, it was hypothesized that a polyphenol oxidase (PPO) catalyzed this step (Steiner et al., 1999). Further work is needed to determine the gene responsible for converting tyrosine to L-DOPA in *Silene*, and to determine how widespread the loss of *CYP76AD1* is in Caryophyllaceae.

Given *Silene latifolia*'s ability to hydroxylate tyrosine, the only betalain synthesis genes it is missing are *DODAa* and *CYP76AD1α.* Since there is no evidence that any of the betalain genes found in *Silene* are non-functional, the loss of these genes may be the reason for its lack of betalain, rather than the loss of *ADHα* as hypothesized. In fact, the similarity in expression of *Silene's ADHβ* to *ADHα* in sea beet, presence of multiple betalain synthesis genes, loss of *FNS*, and non-detection of several Phe-derived metabolites, our expectation that *Silene latifolia* is a reversal to ancestral Phe-dominant metabolism looks unlikely. Instead, *Silene* appears to have lost parts of Phe-derived metabolism, while maintaining some, but not all, elements of Tyr-dominant metabolism.

The relationship between gene expression and metabolite abundance

The abundance of metabolites and expression of genes that are implicated in the production of those metabolites have a complex relationship. In some cases, they align well. For example, none of the *FLS* genes in *Silene* were particularly responsive to any of the treatments and were fairly lowly expressed (Fig 3.9). This is the gene responsible for converting dihydrokaempferol to kaempferol and quercetin, neither of which were detected in *Silene.* Additionally, the low expression of late anthocyanin synthesis genes in both species makes sense in the context of the known lack of anthocyanin in sea beet and the lack of visible pigment accumulation in *Silene.*

Several of the intermediate metabolites in the pigment pathways tend to correlate with pathway genes. For example, in the high light group in both species, p-coumaric acid is positively correlated with members of many of the early anthocyanin pathway genes, including the gene responsible for its production, *C4H* as well as *PAL, CHI* and its immediate successor *4CL*. On the other hand, L-DOPA and dopamine don't appear to correlate more strongly to betalain synthesis genes than they do to shikimate or anthocyanin genes in high light or MeJA treatment groups. In addition, many of the found *MYB* genes correlate strongly with metabolites in both species, suggesting they may play a role in Tyr or Phe pathways (Fig. 3.11).

Part of this can be explained by flux through the pathway, as we are catching a moment in time of a dynamic system at each time point. Rather than accumulation in our measured metabolites, upregulated genes may be moving carbon through to downstream pathways and other branches, and therefore weakens correlations between gene expression vs. intermediate metabolites measured. Branches off the tyrosine pathway are poorly known compared to Phe pathway genes, and so potential correlations between metabolites and genes on that side are not

shown. It is worth noting, however, that the genes we did recover that branch off the pigment pathways don't correlate well either. For example, *HCT*, which is the first step branching off anthocyanin synthesis to produce sinapic acid, is upregulated by high light in both species, but correlates poorly with sinapic acid in the sea beet high light group, and sinapic acid is not detected at all in the *Silene* high light group (Fig. S3.1). One possible explanation, at least for some metabolites, is a time delay between gene expression and metabolite accumulation. Apart from the earlier example of kaempferol and quercetin appearing to be delayed relative to their preceding gene, *CYP76AD1β* expression peaked at Time 1, while L-DOPA peaked at time 3. Other alternative explanations include post-transcriptional regulation of pathway genes decoupling the gene-metabolite relationship, or carbon being pushed into other related metabolites not measured here.

CONCLUSIONS

The relationship between Tyr and Phe metabolism is more complicated than predicted. Rather than the Tyr pathway in betalain-producing sea beet functionally replacing the Phe pathway in anthocyanic *Silene*, the Tyr pathway in sea beet was more responsive to our simulated biotic stress than the Phe pathway, which responded primarily to our high light treatment. While the Tyr pathway seems to share or co-opt some of the Phe pathway regulation, it has been primarily shaped by biotic forces separately from Phe. *Silene latifolia* also displays characteristics of both Tyr- and Phe-enriched metabolism, rather than being a true reversal to Phe-dominant metabolism. Future work should further explore the expression of *ADH* paralogs in Caryophyllales to determine if the expression of *ADHβ* in *Silene* is recovered ancestral *ADHα* function and if *Silene* is unique in its intermediate phenotype among anthocyanic lineages in the core Caryophyllales. Additionally, correlations with the known pathway genes and potential

regulatory genes should be further investigated to understand the shared regulation between Tyr and Phe pathways. Similarly, non-regulatory gene correlations outside of the known pathway genes and untargeted metabolomics may reveal other avenues that phenylalanine and tyrosine are funneled into.

SUPPLEMENTAL MATERIALS

Table S3.1 – Functionally characterized bait loci for pathway reconstruction.

Table S3.2 – List of reconstructed pathway genes found to be differentially expressed.

Figure S3.1 – Expression of putative orthologs to betalain, anthocyanin, shikimate, and branching metabolism pathways.

Table S3.3 – List of differentially expressed genes from the salt experimental treatment and top BLAST hits from *Arabidopsis thaliana*.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

Brockington, S., Y. Yang, F. Gandia-Herrero, S. Covshoff, J.M. Hibberd, R.F. Sage, G.K.S. Wong, M.J. Moore, S.A. Smith. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist* 207: 1170-1180.

Campanella, J.J., J.V. Smalley, M.E. Dempsey. 2014. A phylogenetic examination of the primary anthocyanin production pathway of the Plantae. *Botanical Studies* 55: 10.

Cho, M.-H., O.R.A Corea, H. Yang, D.L. Bedgar, D.D. Laskar, A.M. Anterola, F.A. Moog-Anterola, et al. 2007. Phenylalanine biosynthesis in *Arabidopsis thaliana*: Identification

and characterization of arogenate dehydratases. *Journal of Biological Chemistry* 282: 30827-30835.

Clement, J.S., and T.J. Mabry. 1996. Pigment evolution in the Caryophyllales: a systematic overview. *Bot. Acta* 109: 360-367.

Dohm, J.C., A.E. Minoche, D. Holtgrwe, S., Capella-Gutirrez, G. Zakrzewski, H. Tafer, O. Rupp, T.R. Srensen, R. Stracke, R. Reinhardt. 2014. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature* 505: 546-549.

Dong, N.-Q. and H.-X. Lin. 2021. Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions.

Frese, L. and B. Ford-Lloyd. 2012. Range of Distribution. In *Beta maritima*: *The Origin of Beets*. Springer Nature (Cham, Switzerland).

Gamon, J.A., J. Peñuelas, C.B. Field. 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* 41: 35-44.

Gilman, I.S., J.J. Moreno-Villena, Z.R. Lewis, E.W. Goolsby, E.J. Edwards. 2022. Gene co-expression reveals the modularity and integration C4 and CAM in *Portulaca*. *Plant Physiology* 189: 735-753.

Gitelson, A.A., O.B. Chivkunova, M.N. Merzlyak. 2009. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany* 96: 1861-1868.

Görlach, J., H.R. Raesecke, D. Rentsch, M. Regenass, P. Roy, M. Zala, C. Keel, T. Boller, N. Amrhein, J. Schmid. 1995. *Proc Natl Acad Sci U.S.A* 92: 3166-3170.

Grabherr, M.G., B.J. Hass, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, et al. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644-652.

Hansen, C.C., D.R. Nelson, B.L. Møller, D. Werck-Reichhart. 2021. Plant cytochrome P450 plasticity and evolution. *Molecular Plant* 14: 1244-1265.

Hatlestad, G.J., R.M. Sunnadeniya, N.A. Akhaven, A. Gonzalez, I.L. Goldman, J.M. McGrath. A.M. Lloyd. 2012. The beet *R* locus encodes a new cytochrome P450 required for red betalain production. *Nature Genetics* 44: 816-820.

Hatlestad, G.J., N.A. Akhavan, R.M. Sunnadeniya, L. Elam, S. Cargile, A. Hembad, A. Gonzalez, J.M. McGrath, A.M. Lloyd. 2015. The beet Y locus encodes an anthocyanin

MYB-like protein that activates the betalain red pigment pathway. *Nature Genetics* 47: 92-96.

He, Q., D. Ma, W. Li, L. Xing, H. Zhang, Y. Wang, C. Du, et al. 2023. High quality *Fagopyrum esculentum* genome provides insights into the flavonoid accumulation among different tissues and self-incompatibility. *Journal of Integrative Plant Biology* 65: 1423-1441.

Jain, G. and K.S. Gould. 2015. Are betalain pigments the functional homologues of anthocyanins in plants? *Environmental and Experimental Botany* 119: 48-53.

Katoh, K., D.M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Keith, B., X.N. Dong, F.M. Ausubel, G.R. Fink. 1991. Differential induction of 3-deoxy-D arabino-heptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack. *Proc Natl Acad Sci U.S.A* 88: 8821-8825.

Kurowska, M.M., A. Daszkowska-Golec, M. Gajecka, P. Kościelniak, W. Bierza, I, Szarejko. 2020. Methyl jasmonate affects photosynthesis efficiency, expression of HvTIP Genes, and Nitrogen homeostasis in Barley. *International Journal of Molecular Sciences* 21, 4335.

Langmead, B. and S. Salzberg. 2012, Fast gapped-read alignment with Bowtie2. *Nature Methods* 9: 357-359.

Liu, Y., Y. Tikunov, R.E. Schouten, L.F.M. Marcelis, R.G.F. Visser, A. Bovy, 2018. Anthocyanin biosynthesis and degradation mechanisms in Solanaceous vegetables: A review. *Frontiers in Chemistry* 6: 52.

Lopez-Nieves, S., Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. New Phytologist 217: 896–908.

Love, M.I., W. Huber, S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.

McGrath, J.M., A. Funk, P. Galewski, S. Ou, B. Townsend, K. Davenport, H. Daligault, et al. 2023. A contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.). *DNA Research* 30: 1-14.

Murata, N., Allakhverdiev, S.I., Nishiyama, Y. 2012. The mechanism of photoinhibition *in vivo*:

Re-evaluation of the roles of catalase, a-tocopherol, non-photochemical quenching, and electron transport. *Biochimica et Biophysica Acta - Bioenergetics* 1817: 1127-1133.

Nguyen, L-T., H.A. Schmidt, A. von Haeseler, B.Q. Minh. 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268-274.

Patro, R., G. Duggal, C. Kingsford. 2015. Salmon: Accurate, versatile, and ultrafast quantification from RNA-Seq data using lightweight-alignment. *Biorxiv.*

Piatkowski, B., K. Imwattana, E.A. Tripp, D.J. Weston, A. Healey, J. Schmutz, A.J. Shaw. 2020. Phylogenomics reveals convergent evolution of red-violet coloration in land plants and the origins of the anthocyanic biosynthetic pathway. *Molecular Phylogenetics and Evolution* 151: 106904.

Pucker, B., N. Walker-Hale, J. Dzurlic, W.C. Yim, J.C. Cushman, A. Crum, Y. Yang, S.F. Brockington. 2023. Multiple mechanisms explain loss of anthocyanins from betalain-pigmented Caryophyllales, including repeated wholesale loss of a key anthocyanidin synthesis enzyme. *New Phytologist* 241: 471-489.

Pucker, B. 2022. Automatic identification and annotation of *MYB* gene family members in plants. *BMC Genomics 23:* 220.

Schmieder, R., and R. Edwards. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863-864.

Sheehan, H., T. Feng, N. Walker-Hale, S. Lopez-Nieves, B. Pucker, R. Guo. W.C. Yim, et al. 2020. Evolution of L-DOPA 4,5-dioxygenase activity allows for recurrent specialization to betalain pigmentation in Caryophyllales. *New Phytologist* 227: 914-929.

Soneson, C., M.I. Love, M.D. Robinson. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* 4: 1521.

Steiner, U., W. Schliemann, H. Böhm, D. Strack. 1999. Tyrosinase involved in betalain synthesis of higher plants. *Planta* 208: 114-124.

Stracke, R., D. Holtgräwe, J. Schneider, B. Pucker, T.R. Sörenson, B. Weisshaar. 2014. Genome-wide identification and characterisation of *R2R3-MYB* genes in sugar beet (*Beta vulgaris*). *BMC Plant Biology* 14: 249.

Sunnadeniya, R., A. Bean, M. Brown, N. Akhaven, G. Hatlestad, A. Gonzalez. V.V. Symonds, A. Lloyd. 2016. Tyrosine hydroxylation in betalain pigment biosynthesis is performed by

cytochrome P450 enzymes in beets (*Beta vulgaris*). *PLOS ONE* 11: e0149417.

Timoneda, A., T. Feng, H. Sheehan, N. Walker-Hale, B. Pucker, S. Lopez-Nieves, R. Guo, S. Brockington. 2019. The evolution of betalain biosynthesis in Caryophyllales. *New Phytologist* 224: 71-85.

Tirado, S.B., S.S. Dennis, T.A. Enders, N.M. Springer. 2021. Utilizing spatial variability from hyperspectral imaging to assess variation in maize seedlings. *The Plant Phenome Journal* 4: e20013.

Wang, R., J.A. Gamon, R.A. Montgomery, P.A. Townsend, A.I. Zygielbaum, K. Bitan, D. Tilman, J. Cavendar-Bares. 2016. Seasonal variation in the NDVI-Species Richness relationship in a prairie grassland experiment (Cedar Creek). *Remote Sensing* 8: 128.

Winkler, T.S., S.K. Vollmer, N. Dyballa-Rukes, S. Metzger, M.G. Stetter. 2024. Isoform-resolved genome annotation enables mapping of tissue-specific betalain regulation in amaranth. *New Phytologist* 243: 1082-1100.

Xie, F., C. Chen, J. Chen, J. Chen, Q. Hua, K. Shah, Z. Zhang, et al. 2023. Betalain biosynthesis in red pulp pitaya is regulated via HuMYB132: a R-R type MYB transcription factor. *BMC Plant Biology* 23: 28.

Yokoyama, R., M.V.V. de Oliveira, B. Kleven, H.A. Maeda. 2021. The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation. *The Plant Cell* 33: 671-696.

Yue, J. M. Krasovec, Y. Kazama, X. Zhang, W. Xie, S. Zhang, X. Xu, et al. 2023. The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome. *Current Biology* 33: 2504-2514.e3.

CHAPTER 4: UNTARGETED METABOLOMICS IN CARYOPHYLLALES EXHIBIT LINEAGE-SPECIFIC METABOLIC PROFILES

INTRODUCTION:

Plant specialized metabolism serves a number of roles in the plant, including enabling interactions with other organisms and protection from abiotic factors (Weng et al., 2021). For example, flavonoids and other phenylpropanoids are known to provide photoprotection and across the plant kingdom are induced by high light intensity (Lingwan et al., 2023). Similarly, a wide variety of defensive compounds, including phenylpropanoids, are known to be induced by herbivory via signaling by jasmonic acid/methyl jasmonate (Wasternack and Hause, 2013). Specialized metabolite (SM) pathways branch off primary metabolism. For example, the phenylpropanoid class of SMs are synthesized from the amino acid phenylalanine (and tyrosine in some lineages; Barros et al., 2016). Due to the complexity and expansiveness of primary metabolism, and the potential for further branching and modification of SM pathways, SM is exceedingly diverse across the plant kingdom, potentially ranging up to 1,000,000 different metabolites (Afendi et al. 2012). Some SM classes are widespread across the plant kingdom, while others more lineage specific. For example, the anthocyanin pigments, which belong to the larger phenylpropanoid compound class, are nearly ubiquitous throughout the plant kingdom. On the other hand, glucosinolates, a class of compounds synthesized from glucose and amino acids, is restricted to the order Brassicales and the family Putranjivaceae (Rodman et al., 1998). Several SMs have also evolved independently in different lineages, such as caffeine, which occurs in several families, and seems to have evolved via two different biosynthetic pathways (Huang et al., 2016).

The complexity and diversity of specialized metabolism, even within a lineage, makes SMs challenging to study. Although there is phylogenetic signal in specialized metabolism, where closely related plants tend to make the same or similar metabolites, diversification of a SM pathway can evolve quickly and the metabolomic profiles of closely related species may have major or minor divergences (Holeski et al., 2021; Lichman et al., 2020; Youssef et al., 2023). In addition, as there is a resource cost in SM production, many SMs may not be constantly produced or produced on low amounts under "normal" conditions, instead being synthesized only in specific conditions where their presence benefits the plant (Tiedge et al., 2022; Jeon et al., 2018; Moroldo et al., 2024).

Further, a relatively small amount of plant specialized metabolism is known. The Natural Product Activity & Species Source Database (NPASS), which compiles known plant SMs, currently has entries for ~96,000 metabolites across ~32,000 species, including bacteria and fungi as well as plants (Zhao et al., 2023). This is a far cry from the estimated 1,000,000 existing metabolites across the plant kingdom. Screening more plant species for more SMs is time and resource costly. While methods such as liquid or gas chromatography-mass spectrometry (LC/GC-MS) allow us to observe a wide swath of the SMs present in a species, no one protocol will capture the full SM diversity in a sample. Additionally, moving from a peak spectral data to a structurally characterized metabolite requires further analysis of the compound or comparison to limited existing public data, making it logistically intractable to characterize a species full metabolome (Moghe and Kruse, 2018). Even when a metabolite is observed and characterized, biosynthetic pathway elucidation presents its own challenges. The same metabolites may not be synthesized via the same pathway across the plant kingdom, as in caffeine. Gene

duplication/neofunctionalization and enzyme promiscuity are also crucial features contributing to the evolution of metabolite diversity (Moghe and Kruse, 2018).

The plant order Caryophyllales is a compelling system to better understand specialized metabolism. Containing ~12,500 species across 37 families, the order is home to important crop species such as *Beta vulgaris* L. (beets), *Fagopyrum* spp. (buckwheat), and *Chenopodium quinoa* Willd. (quinoa); well-known medicinal plants like *Simmondsia chinensis* (Link) C.K.Schneid. (Jojoba) and *Lophophora williamsii* (Lem. ex Salm-Dyck) J.M.Coult. (peyote); and popular horticultural species including *Dianthus* spp. and *Celosia* spp. Caryophyllales is a notable group for the wide variety of extreme habitats its members have adapted to live in, from deserts to salt marshes to gypsum soil to nutrient-poor bogs.

Caryophyllales is also notable for a diversification of specialized metabolites derived from the amino acid tyrosine (Lopez-Nieves et al., 2018). This diversification was enabled by the duplication of the gene arogenate dehydrogenase (*ADH*) in Caryophyllales (Fig 3.1**)**. The *ADH* gene encodes an enzyme responsible for converting the output of the primary metabolism shikimate pathway, arogenate, into tyrosine (Tyr). While the canonical copy of *ADH* (*ADHβ),* is feedback inhibited, the core-Caryophyllales specific-copy (*ADHa)* has relaxed feedback inhibition, allowing for increased Tyr production (Lopez-Nieves et al., 2018). As a result of this increased availability of Tyr, lineages in the core Caryophyllales appear to have a Tyr-enriched specialized metabolism. This is exemplified by the betalain pigments, which are a unique class of pigments found exclusively in the core Caryophyllales (at least in the plant kingdom-one fungal lineage has also independently evolved betalain pigments; Clement and Mabry 1996; Gill and Steglich 1987). However, other tyrosine derived metabolites also are enriched in the order, including catecholamines, isoquinoline alkaloids, etc (Busta et al., 2024; Watkins et al. 2023).

Besides tyrosine, arogenate is also a precursor of phenylalanine (Phe), suggesting that increased availability of Tyr in Caryophyllales may come at the cost of reduced availability of Phe (Cho et al., 2007). One piece of evidence for this is the loss of Phe-derived anthocyanin pigments in lineages that have developed betalains, which originated multiple times in Caryophyllales (Fig 3.1; Clement and Mabry 1996; Lopez-Nieves et al., 2018). Anthocyanin-producing lineages in the core Caryophyllales, on the other hand, lose their *ADHa*, although basal members of these lineages may retain a functional or pseudogenized copy (Lopez-Nieves et al., 2018).

This duplication and repeated losses of *ADHa* may have set up a pattern of switching between Tyr- or Phe-dominant metabolism in the order, with the non-core Caryophyllales maintaining the ancestral Phe-dominant metabolism, Tyr-dominant metabolism in the core, with reversals to to Phe-dominant metabolism with the loss of *ADHa* . However, it is unknown whether the reversal lineages are true reversals back to the ancestral Phe metabolism. Additionally, the consequences of the gain and loss of *ADHa* on overall specialized metabolism in Caryophyllales is poorly understood outside of a few targeted SMs (Busta et al., 2024).

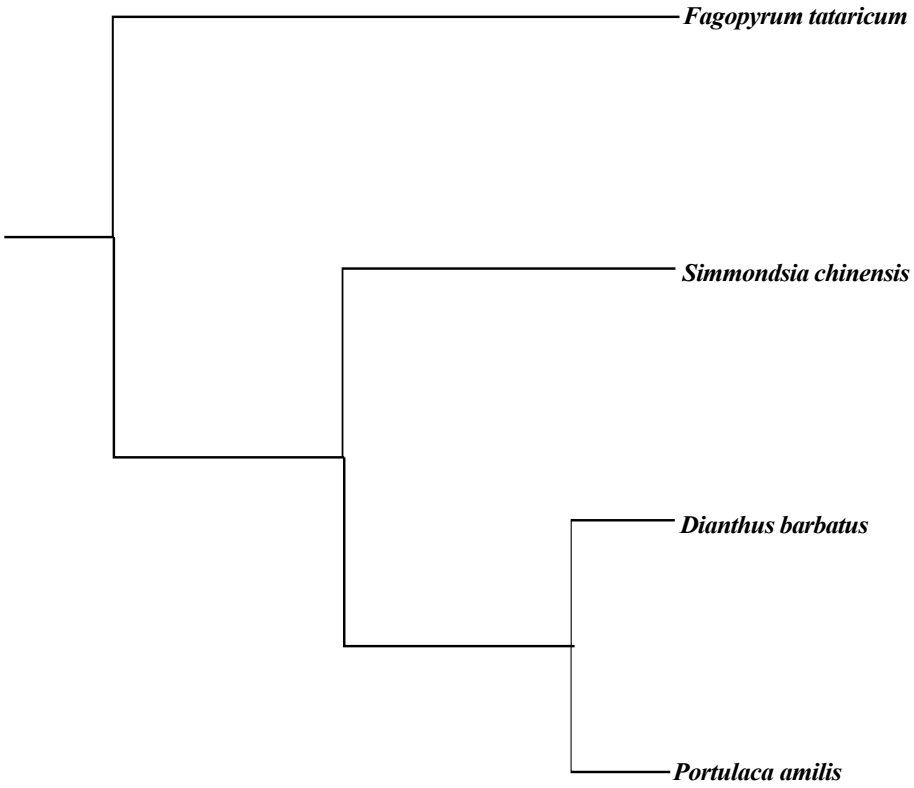
Although the gain and loss of a gene with such major consequences on amino acid accumulation is already interesting, its impacts on downstream specialized metabolism, is especially interesting to investigate, as the phenylpropanoid pathway from Phe includes SMs such as lignins, flavonoids, some types of tannins etc. Many of these phenylpropanoids are common throughout the plant kingdom and are known for playing major roles in plant defense, protection from abiotics, and even structure (Landis et al. 2015; Liu et al., 2018; Sharma et al. 2019). The impacts of siphoning precursors away from this pathway would be expected to

impact the ways Tyr-dominant lineages deal with stressors, including compensating for a lack of phenylpropanoid abundance and diversity with tyrosine-derived metabolites.

To explore Phe- and Tyr-derived metabolism in the order, here we focus on four selected species, each representing a lineage of a specific metabolism type (Fig 4.1). *Fagopyrum tataricum* (L.) Gaertn. (Polygonaceae) is a semi-domesticated member of the non-core clade diverging before the duplication of *ADH,* and represents the ancestral Phe-dominant metabolism. *Simmondsia chinensis* (Link) C.K.Schneid (Simmondsiaceae) is an early-diverging member of the core clade that lives in arid climates of North America , and has a copy of *ADHa* that may be functional or pseudogenized. It is unknown whether *S. chinensis* has Tyr or Phe-dominant metabolism. *Dianthus barbatus* L. (Caryophyllaceae) is native to Europe and Asia and is an anthocyanin producing member of the core clade apparently lacking *ADHa,* and is expected to be a reversal back to Phe-dominant metabolism. Lastly, *Portulaca amilis* Speg. (Portulacaceae) is native to South America and has been introduced to disturbed areas in the U.S. It is a betalain-producing member of the core clade with a copy of *ADHa,* and represents the Tyr-dominant metabolism type.

We expected that the lineages without a working copy of *ADHa* would rely more on phenylpropanoid SMs to mediate stressful conditions, while the Tyr-dominant lineage compensates for these conditions with more Tyr-derived SMs. To test this prediction, as well as explore overall SM diversity in these species, we treated plants of each focal species with high light and simulated herbivory via methyl jasmonate, and used untargeted metabolomics to explore the differences between species and groups.

Fig 4.1-A simple cladogram of the four focal species. *Fagopyrum tataricum* and *D. barbatus* do not have *ADHa,* and are expected to be Phe-dominant. *Simmondsia chinensis* and *P. amilis* have *ADHa*, and are expected to be a transitional metabolism type and Tyr-dominant, respectively.



MATERIALS AND METHODS

Seed germination and growth

Seeds of *F. tataricum* were obtained from USDA-GRIN (PI 673875). Seeds from *P. amilis* came from the same line as used in Gilman et al. (2022). *Dianthus barbatus* seeds were wild collected from a plant at Reservoir Woods in Roseville, MN, USA (45.002730, -93.135302; Crum 55, MIN). *Simmondsia chinensis* seeds were purchased from Strictly Medicinal Seeds, LLC (Williams, Oregon, USA). *Simmondsia chinensis* and *P. amilis* did not germinate as readily as *D. barbatus* and *F. tataricum*, and were treated with a 15-minute soak in a 15% bleach and 0.015% Silwet solution before planting to break dormancy. *Dianthus barbatus* and *F. tataricum* were sown directly without pre-treatment. All seeds were germinated in trays of SunGro Sunshine #1

potting mix (Agawam, MA, USA), before being transplanted to ⅜ x 5 inch Anderson Band Bottom Pots (Portland, Oregon, USA) with the same type of potting mix and 1 gram of Osmocote 14-14-14 fertilizer. Plants were grown in a growth chamber ~200 mmol/m^2/s light intensity. Growth chambers were set on a 14-hour daytime schedule with 70% humidity and a daytime temperature of 23 °C. Temperature fell to 20 °C at night. Temperature changes were ramped up or down over a period of 30 minutes at the start and end of the daytime period. Plants were watered uniformly 3 times a week. Experiments began when at least 28 individuals of a species had at least three fully expanded leaf pairs or four fully expanded leaves for species with alternate leaf arrangement.

Experimental design

Within each species, plants were divided into three treatment groups. Control treatment group plants were kept in the same growth chamber they were germinated in. The plants treated with methyl jasmonate were dipped into a petri dish with a 0.5 mM methyl jasmonate (MeJA) solution, prepared with 0.015% Silwet. L-77 (Fisher Scientific), and 0.25% 190 proof ethanol, making sure to fully immerse the leaves designated for sampling. MeJA-treated plants were placed back in the same chamber as the control group after allowing the volatile MeJA to disperse for 15 minutes. High light treatment plants were placed into a growth chamber with ~900 mmol/m^2/s light intensity at the start of the experiment, keeping the rest of the chamber settings the same as the control chamber. All treatments were applied and the experiment initiated at midday in the growth chamber.

Tissue sampling occurred at three timepoints. A sub-group of four biological replicates in the control group were sampled at the start of the experiment (Time 0, midday on day 1). At the

following two timepoints, four replicates from each group were sampled, for a total of 12 control group plants and 8 each in the methyl jasmonate and high light groups (Table 4.1). Sampling of each plant consisted of flash freezing in liquid nitrogen 0.2 grams of leaf tissue from the second fully expanded leaf pair when possible. Some leaves, particularly in *D. barbatus* were smaller and as little as 0.05grams tissue were collected. When leaf arrangement was alternate, the second oldest out of the four fully expanded true leaf was used for chemical extraction for untargeted UHPLC-MS, and the third oldest was stored in a -80 °C freezer for potential future RNA-sequencing. When leaf arrangement was opposite, no discrimination was made between the two leaves of the second oldest leaf pair.

Table 4.1-Replicates sampled at each timepoint in each treatment

Treatment	**Time 0 (0 hours)**	**Time 1 (48 hours)**	**Time 2 (flexible)**
Control	4 plants sampled	4 plants sampled	4 plants sampled
MeJA	0 plants sampled	4 plants sampled	4 plants sampled
High Light	0 plants sampled	4 plants sampled	4 plants sampled

The first timepoint where treatment groups were sampled (Time 1) occurred at 48 hours after treatments were applied. The second and final time point (Time 2) was determined by leaf reflectance spectra. At each timepoint, and in between Time 1 and Time 2, leaf reflectance of the oldest fully expanded leaf was measured using a BLUE-Wave spectrometer with a 300-1100 nm range (StellarNet Inc, Tampa, FL, USA). Because different plant species have different rates and intensities of stress responses, the leaf reflectance was used to determine the endpoint of the experiment for each species. Tissue sampling for the final time point occurred when the measured raw leaf reflectance of the high light treatment group visually differentiated from the control group in a PCA plot generated in R. Previous experimentation showed that MeJA

treatment does not result in significant changes to leaf spectral indices, so only high light was used to make the determination here (Chapter 3). In addition, leaf reflectance was measured one week after the end of the experiment to verify plant recovery.

Chemical extraction and UHPLC-MS

Metabolites were extracted from leaf tissue using 80% methanol. For every 0.2 g of sample mass, 1 mL of methanol was added to the sample in a 1.5 mL microcentrifuge tube. The leaf tissue samples with methanol added were pickled in a -80 C freezer for one week. The samples were then removed from the freezer, and a tungsten bead was added to each tube. Samples were then homogenized using a Spex SamplePrep model 2010 Geno/Grinder for 5-10 minutes (depending on leaf toughness) at 1500 rpm. After centrifugation at 14000 rpm for 5 minutes, the supernatant of each sample was transferred to a new tube and stored at -80 C until analysis.

Untargeted metabolomics was carried out on an ultra-high performance liquid chromatographer (UHPLC) - quadropole orbitrap mass spectrometer (Ultimate® 3000 HPLC, Q Exactive™, Thermo Fisher Scientific, Waltham, MA, USA) with an autosampler and 4 C sample vial block. Chromatographic separations were carried out using an Acquity C18 HSS T3 1.8 um particle size, 2.1 x 100 mm (Waters, Milford, MA, USA) column at 40 C. Flow rate was set to 0.4 mL/min. Samples were run on a 24-minute gradient using the mobile phases A: 0.1% formic acid in water and B: 0.1% formic acid in acetonitrile and the following gradient elution profile: initial 2% B, 2 min 2% B, 20 min 98% B, 1 min 98% B, 1 min 2% B. Injection volumes were 1 uL per sample. Mass spectrometer (MS) conditions included a full mass scan range of 125-1800 m/z, resolution 70,000, desolvation temperature 350 C, spray voltage 3800 V, auxiliary gas flow rate 50, sheath gas flow rate 50, sweep gas flow rate 1, S-Lens RF level 50, and auxiliary gas

heater temperature 300 C. Samples were randomized within species, with water blanks run in between species and 5 times throughout a species block.Xcalibur software v2.1 was used for data collection and visualization (Thermo Fisher Scientific, Waltham, MA, USA).

Data processing and analysis

Raw UHPLC-MS data files were converted into mzML format using the tool msconvert by ProteoWizard (Chambers et al. 2012). Data was processed in MzMine v 2.53 (Pluskal et al., 2010). Positive ion mode and negative ion mode were processed separately. Data was processed in batches by species, with noise levels set between 2E5 and 5E5, and retention time allowance between 0.04 mins and 0.1 mins. The resulting chromatograms were aligned together, and exported to a csv file containing both positive and negative ion mode features. Results were checked manually for duplicate features.

Data was log-transformed (base 2) and differences between treatments and species were tested for using partial least squares discrimination analysis (PLS-DA) and orthogonal partial least squares discrimination analysis (OPLS-DA), using the ropls package in R. Confusion matrices for OPLS-DA models were trained by partitioning the dataset into two equal subsets, with one as the training set and the other as the test subset.

RESULTS

Spectral results

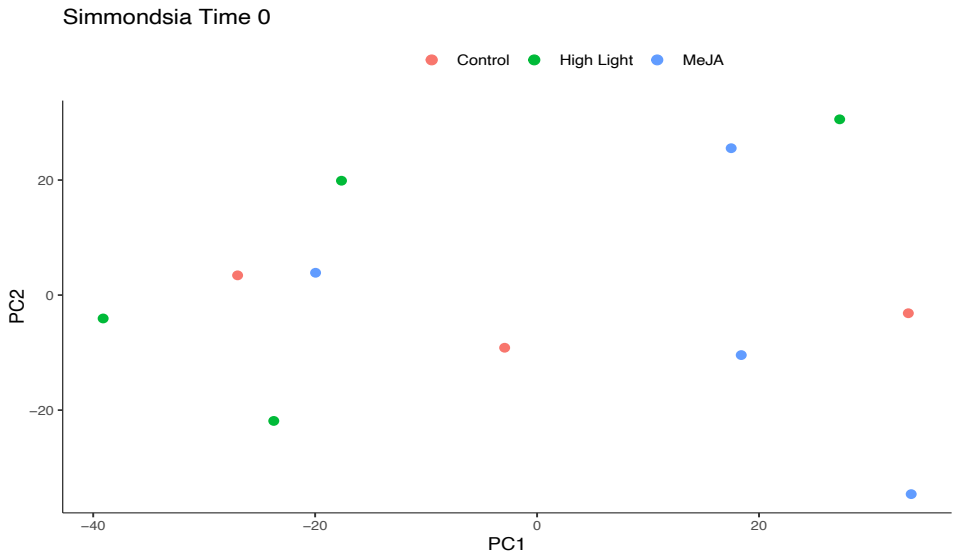
Leaf reflectance spectra between plants in the control treatment and plants in the high light treatment clustered separately across all species (Fig 4.2). However, the amount of time before the high light and control separated and the experiment ended was variable among species (Table 4.2). The MeJA group plants never visually separately from the controls. Similarly, visually, the

high light plants were noticeably different than control plants, but MeJA-treated plants were not (Fig 4.3). In addition, the scans taken one-week post-experiment showed the treatment groups coming back together, indicating recovery (Fig. 4.2).

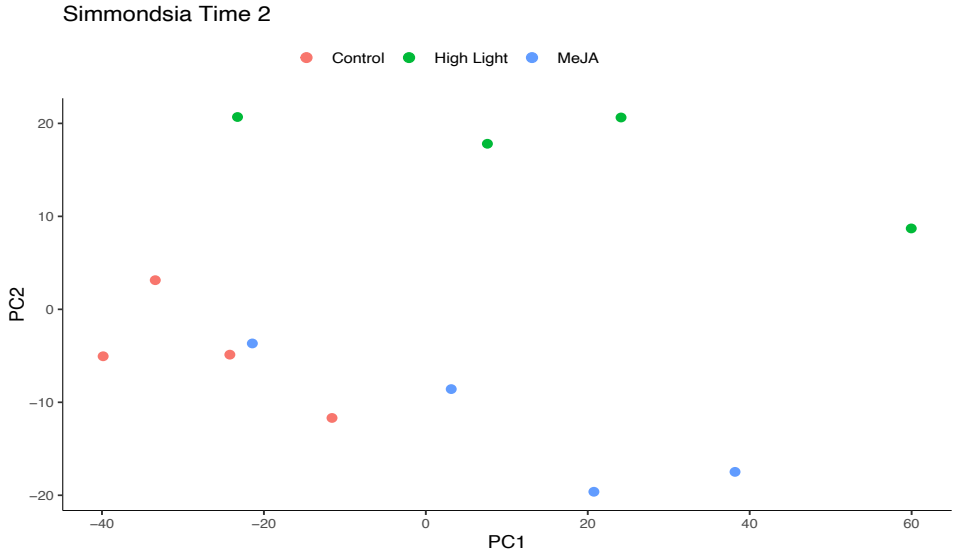
Table 4.2- Number of days the experiment ran for each species. The experiment ran until the control and high light group plants visually separated on a PCA of the leaf reflectance scans, at which point leaves were sampled for Time 2.

Species	**Number of days to Time 2**
Fagopyrum tataricum	6
Simmondsia chinensis	22
Dianthus barbatus	13
Portulaca amilis	6

a.)


Simmondsia Time 0
Control
High Light
MeJA
PC2
PC1
20
0
−20
−40
−20
0
20

b.)


Simmondsia Time 2
Control
High Light
MeJA
PC2
PC1
20
10
0
−10
−20
−40
−20
0
20
40
60

c.)

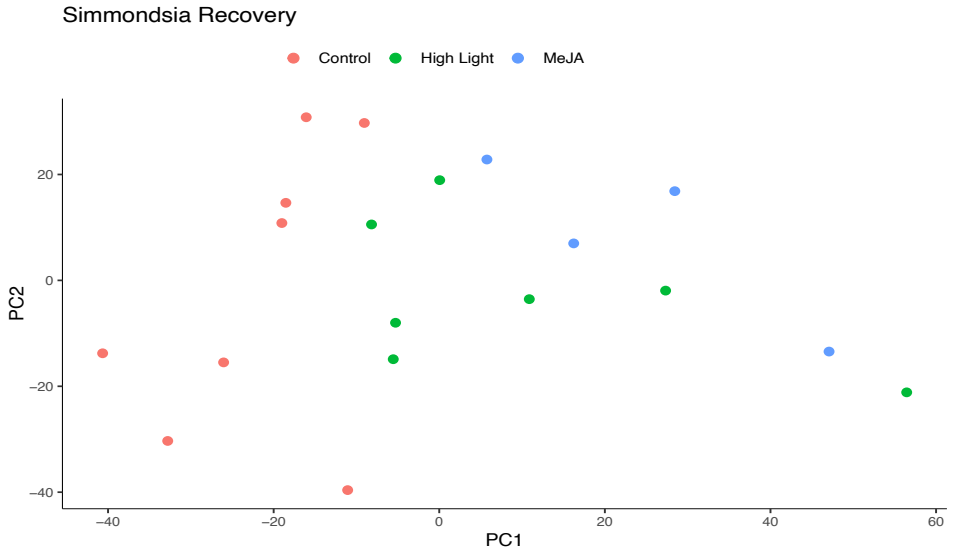

Simmondsia Recovery
Control
High Light
MeJA
PC2
PC1
20
0
−20
−40
−40
−20
0
20
40
60

Fig. 4.2-An example of PCAs generated with non-normalized leaf reflectance data for *Simmondsia*. a.) Time 0 with a random distribution. b.) Time 2 (22 days). High light plants are separated from Control and MeJA-treated plants c.) Recovery scan 7 days after Time 2. High Light-treated plants have come back together with the other plants.

Fig. 4.3- As an example**,** *Dianthus barbatus* Control, MeJA, and High Light treatment groups at Time 2, highlighting the lack of visible pigment accumulation in the MeJA group and yellowing of the High Light group.

Untargeted metabolomics between species and treatments

The total number of features included in analysis across species, treatments, and ion mode was 2537. The total number of samples came to 112 (28 samples per species for four species). PCA analysis of all samples showed a clear discrimination of species, with PC1

explaining 39% variation and PC2 explaining 23% (Fig 4.4). A heat map visualization of features present in each sample also showed a clear difference by species (Fig. 4.5). A PLS-DA model built to discriminate by species was significant (R^2Y=0.992). PLS-DA analyses of treatment or timepoint across species were not significant, so further analyses were conducted within each species separately. In all species, an OPLS-DA model comparing the MeJA groups to the control were insignificant or unable to be built due to the first component being insignificant. These groups also were undistinguishable in a PCA (Fig 4.4). In addition, models built discriminating within species based on timepoint were insignificant.

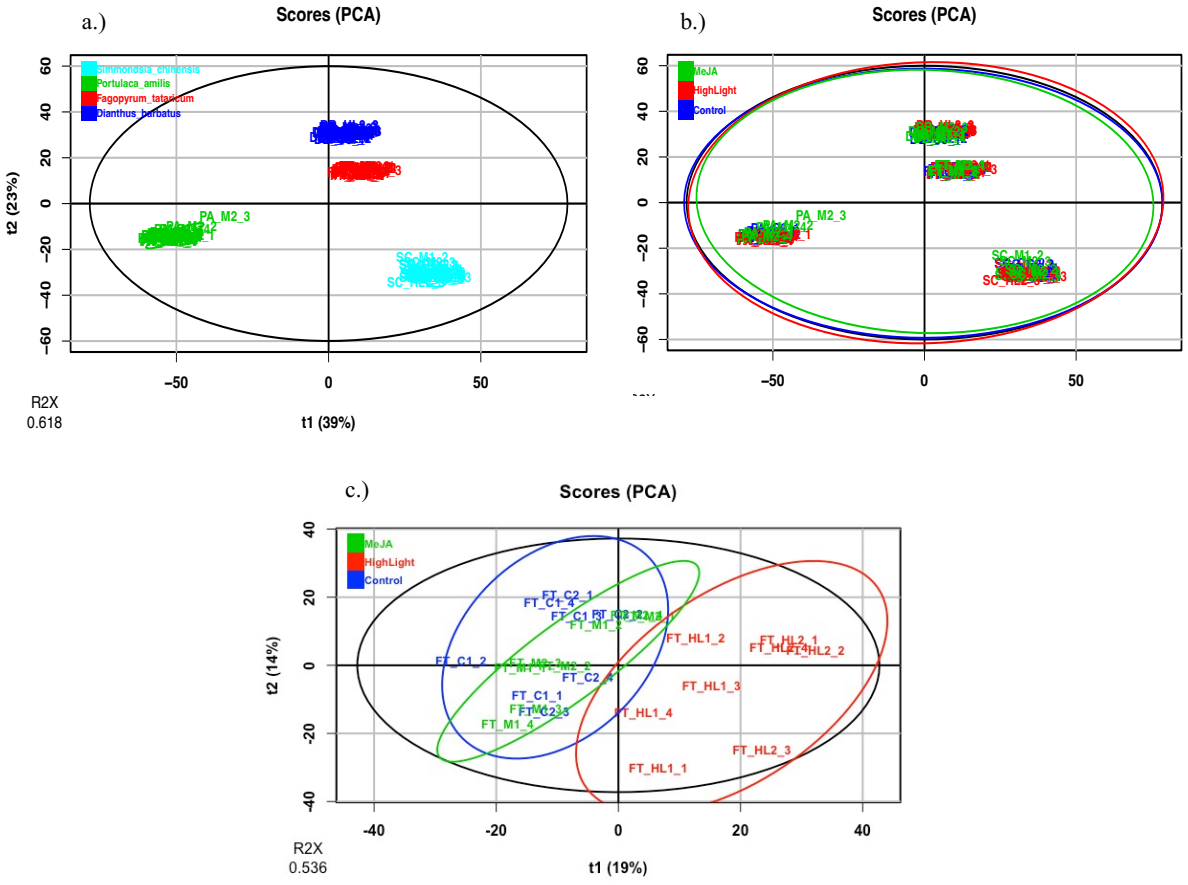


Fig 4.4- PCAs of all species and treatments showing a.) distinct clustering of samples based on species and b.) no discernible differentiation by treatment and c.) An example PCA showing *F. tataricum* samples at Times 1 and 2. The high light plants of the two timepoints form a distinct group, potentially drowning out metabolite differences by timepoint. The x and y axes depict principal components 1 and 2, respectively. The black ellipsis represents the global confidence limit, and colored ellipses represent the confidence limits of the group.

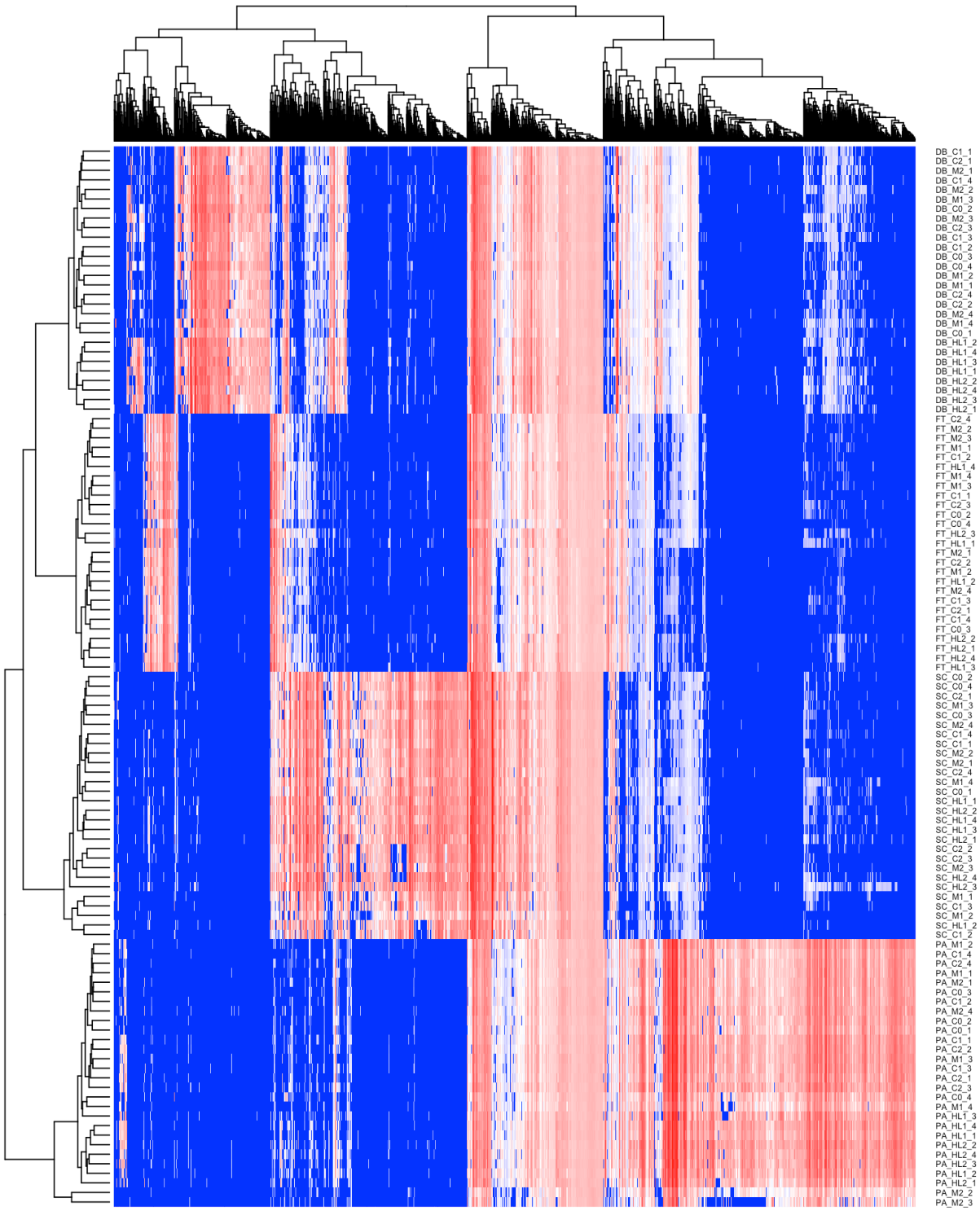

DB_C1_1
DB_C2_1
DB_M2_1
DB_C1_4
DB_M2_2
DB_M1_3
DB_C0_2
DB_M2_3
DB_C2_3
DB_C1_3
DB_C1_2
DB_C0_3
DB_C0_4
DB_M1_2
DB_M1_1
DB_C2_4
DB_C2_2
DB_M2_4
DB_M1_4
DB_C0_1
DB_HL1_2
DB_HL1_4
DB_HL1_3
DB_HL1_1
DB_HL2_2
DB_HL2_4
DB_HL2_3
DB_HL2_1
FT_C2_4
FT_M2_2
FT_M2_3
FT_M1_1
FT_C1_2
FT_HL1_4
FT_M1_4
FT_M1_3
FT_C1_1
FT_C2_3
FT_C0_2
FT_C0_4
FT_HL2_3
FT_HL1_1
FT_M2_1
FT_C2_2
FT_M1_2
FT_HL1_2
FT_M2_4
FT_C1_3
FT_C2_1
FT_C1_4
FT_C0_3
FT_HL2_2
FT_HL2_1
FT_HL2_4
FT_HL1_3
SC_C0_2
SC_C0_4
SC_C2_1
SC_M1_3
SC_C0_3
SC_M2_4
SC_C1_4
SC_C1_1
SC_M2_2
SC_M2_1
SC_C2_4
SC_M1_4
SC_C0_1
SC_HL1_1
SC_HL2_2
SC_HL1_4
SC_HL1_3
SC_HL2_1
SC_C2_2
SC_C2_3
SC_M2_3
SC_HL2_4
SC_HL2_3
SC_M1_1
SC_C1_3
SC_M1_2
SC_HL1_2
SC_C1_2
PA_M1_2
PA_C1_4
PA_C2_4
PA_M1_1
PA_M2_1
PA_C0_3
PA_C1_2
PA_M2_4
PA_C0_2
PA_C0_1
PA_C1_1
PA_C2_2
PA_M1_3
PA_C1_3
PA_C2_1
PA_C2_3
PA_C0_4
PA_M1_4
PA_HL1_3
PA_HL1_4
PA_HL1_1
PA_HL2_2
PA_HL2_4
PA_HL2_3
PA_HL1_2
PA_HL2_1
PA_M2_2
PA_M2_3

Fig. 4.5-A heat map of all features from untargeted UHPLC-MS. Rows are samples, and columns LC-MS features. Samples are designated in the format of Species_Treatment/Time_Replicate, where DB= *Dianthus barbatus,* FT= *Fagopyrum tataricum,* SC=*Simmondsia chinensis,* PA= *Portulaca amilis,* C=Control, M=MeJA, and HL= High Light. Color indicates peak height.

In all four species, the model discriminating between high light and control plants performed well, although pR^2Y in *S. chinensis* was > 0.05 (Table 4.3; Fig 4.6). In all four species the OPLS-DA high light: control model predicted the treatments of the test set with good accuracy. No control group samples were misassigned, and with the exception of *D. barbatus*, which predicted group membership of the test set perfectly, all other species consistently assigned one out of four high light group members to the control for an overall success rate of 87%.

In addition, Variable Importance in Projections (VIP) scores were calculated for each High Light:Control model to determine which metabolite features were driving the model. The *F. tataricum* High Light:Control model had 449 VIPs 1. The *S. chinensis* model had 463 features with VIPS 1. The *D. barbatus* model had 511 VIPs 1, while the *P. amilis* model had 563. In pairwise comparisons *D, barbatus* and *F. tataricum* shared the most features with significant VIP scores (181) and *P. amilis* and *S. chinensis* shared the fewest (142). 36 features had VIP scores 1 across all four models, indicating that they drive variation between the high light and control groups in every species (Table 4.5).

Table 4.3- Model statistics for OPLS-DA models built comparing high light samples and control samples. Models were built separately for each species. The R^2Y statistic indicates the degree the classifier (treatment group) explains the variation in the model's predictive component, while Q^2Y is a measure of model fit.

Species	R^2Y	pR^2Y	Q^2Y	pQ^2Y
F. tataricum	0.96	0.05	0.815	0.05
S. chinensis	0.998	0.3	0.669	0.05
D. barbatus	0.999	0.05	0.782	0.05
P. amilis	0.993	0.05	0.82	0.05

Table 4.4- Number of shared features significantly contributing to separating high light and control groups in pairwise comparisons of all four species.

Pairwise Comparison	Shared features with VIPs 1
D. barbatus : F. tataricum	181
D. barbatus : S. chinensis	146
D. barbatus : P. amilis	177
F. tataricum : S. chinensis	144
F. tataricum : P. amilis	171
S. chinensis : P. amilis	142

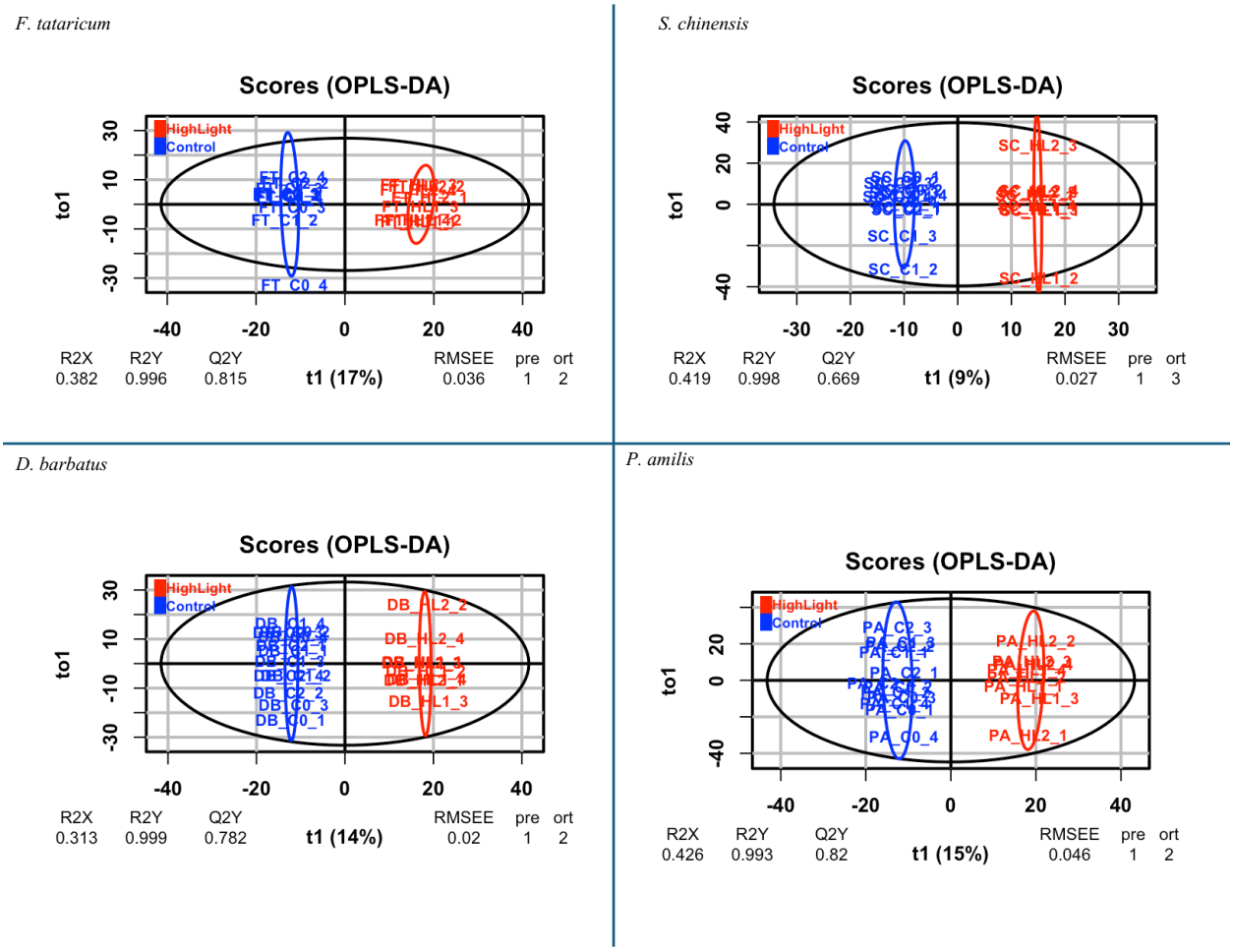


Fig. 4.6- Clustering by OPLS-DA model differentiating high light and control groups in each species. Here, the x axis represents the between group variation (the predictive component of the model), while the y axis represents systemic variation (the orthogonal component). Like the PCA, ellipses represent confidence limits.

Table 4.5- Shared features across all four species with VIPS 1. Ion mode, m/z, and retention time (in minutes) are given.

Ion Mode	m/z	Retention Time
Negative	128.0356	1.30
Negative	283.1043	1.38
Negative	309.0473	0.80
Negative	325.0940	6.74
Negative	331.0682	2.20
Negative	341.0888	6.36
Negative	341.1098	0.76
Negative	362.9670	0.88
Negative	377.0867	0.76
Negative	379.0834	0.76
Negative	387.1157	0.76
Negative	388.1190	0.76
Negative	404.1055	0.76
Negative	415.1989	8.67
Negative	417.2147	8.42
Negative	429.1781	8.93
Negative	431.1054	0.76
Negative	431.1936	7.64
Negative	433.2093	7.01
Negative	433.2094	8.22
Negative	433.2095	7.56
Negative	447.0951	8.28
Negative	465.1052	6.70
Negative	481.2587	16.89
Negative	483.2744	17.38
Negative	549.1690	0.77
Negative	609.1484	8.03
Negative	609.1489	7.81
Negative	611.1647	6.51

Negative	657.2590	12.10
Negative	683.2277	0.76
Negative	315. 0732	4.11
Positive	166.0485	0.70
Positive	343.1212	0.76
Positive	355.1026	5.01
Positive	685.2332	0.76

DISCUSSION

Utility of spectra as an effective tool for standardizing sampling timepoints across species

Making meaningful comparisons among species in stress time course with stress treatments are difficult, as the different species respond differently to stresses at the transcriptional and phenotypical level (Meng et al., 2020; Zhang et al., 2022). Therefore, the same intensity and duration of treatment does not result in comparable levels of response by different plant species. Here, we utilized leaf reflectance to determine if a response to treatments were detectable at the phenotypic level. While the clustering of different groups based on a PCA did not provide a direct quantitative measure on the degree of stress a plant was experiencing, it did allow us to verify that there was a detectable response affecting the leaf physiology. As would be expected for plants living in a wide variety of habitats, the time it took for the high light groups to separate from controls varied, suggesting varied stress responses. *Simmondsia chinensis* took the longest to differentiate, which was expected given that it is the only desert species and has the toughest leaves. After a week after time point 2, the plants coming back together in the PCA suggests a return to similar physiology (Table 4.2). Leaf reflectance may have utility to standardize multi-species light-related stress experiments in the future.

On the other hand, the lack of a response in the spectra for the plants treated with MeJA could be due to a few potential factors. The results of the previous chapter suggest that the

induction of genes associated with responding to methyl jasmonate have a faster response than those responding to high light. It is possible that, with the first timepoint being 48 hours after treatment, the MeJA response has come and gone. Alternatively, the response to MeJA may not involve changing leaf physiology or the metabolites present in a way that is captured at the wavelengths measured, which include the visible and part of the near infrared spectra. At these wavelengths notable features in the spectra include pigments such as chlorophyll, anthocyanin, and betalains. Visibly, no pigment change was observed in the MeJA-treated groups over the time course (Fig 4.2). However, other experiments have observed pigment accumulation or loss in response to MeJA, albeit at higher concentrations in other species (Kim et al., 2011; Milech et al., 2023). Finally, the lack of response to MeJA in the spectra could indicate that there was no actual response to treatment. This is supported by the lack of a significant OPLS-DA model that differentiates the control and MeJA treatment group in any species.

Overall metabolomic differences between species

All four species differentiated on an unsupervised PCA and in a supervised PLS-DA model. Notably, the two species that do not have a copy of *ADHa, F. tataricum* and *D. barbatus,* were nearer to each other on the PCA than the other two species, indicating greater similarity between the two. These species also were sister to each other in cluster analyses (Fig. 4.4). This supports our hypothesis that, at least in terms of metabolite profile, *D. barbatus* is an evolutionary reversal back to pre-*ADHa* metabolism.

The two species with a copy of *ADHa, S. chinensis* and *P. amilis*, were roughly as distant from each other as either was to the two non-*ADHa* species. This, despite *S. chinensis* actually being a closer relative to *D. barbatus* than to *F. tataricum,* implies that whether or not the copy of *ADHa* in *S. chinensis* is still functional, its presence has influenced *S. chinensis* specialized

metabolism to differentiate it from species that either never had or whose ancestors lost the gene entirely. However, it may not have fully developed Tyr-dominant metabolism. Its metabolite profile is distant from that of the TYr-dominant *Portulaca.* Also, in hierarchical clustering, the *S. chinensis* samples did cluster sister to the *Fagopyrum/Dianthus* clade, indicating more similarity with them than with the sole betalain producing species, *Portulaca amilis,* and supporting its status as a transitional species between Phe and Tyr-dominant metabolism.

Metabolomic responses within species in response to time

Within each species, there was OPLS-DA was unable to distinguish samples based on timepoint. Comparisons between Time 0 and the other timepoints are biased by unbalanced groups, as there are only 4 replicates at Time 0, all in the control group, while the other two timepoints include 12 plants, including the meJA and high light treatment groups (Brereton and Lloyd, 2014). Filtering further to only compare the same treatment at different timepoints also resulted in insignificant models, potentially due to the decrease in statistical power that accompanies smaller sample sizes in multivariate data, as each group consisted of only 4 plants in this analysis (Saccenti and Timmerman, 2016).

Comparisons between Time 1 and 2 should not have suffered from balance issues and had comparable sample size to models built on treatment within species. Our inability to distinguish between the two timepoints may be because the overall effect of high light treatment results in within group variation drowning out variation between time points, particularly within the Time 2 group. This is supported by PCAs generated using within-species data at only these two time points, where the high light plants form a distinct group from the other plants at these time points (Fig. 4.4).

Metabolomic responses within species in response to treatment

A significant OPLS-DA model was able to differentiate high light and control groups across all 4 species. The high light groups were usually visually different, accumulating pigment, but also appearing less green and shorter (Fig 4.3). While all four species had a definite high light response, the specialized metabolomic component of those responses were evidently very different between species, given that only 36 spectra features were significantly contributing to differentiating high light groups from the control. Although pairwise VIP comparisons overlapped more, less than half of the most responsive metabolites in each species were shared with another species, indicating a huge diversity of responses to high light across species and suggesting that the different lineages have evolved different strategies to cope with high light stress.

While this variation is certainly partially driven by other pathways besides Tyr and Phe metabolism, the similarity between *Fagopyrum* and *Dianthus* suggests that the Tyr and Phe pathways are playing a part. The Phe pathway especially is known to be inducible with high light stress (Lingwan et al. 2023). It is telling that the species that consistently overlaps the least with the other three is *S. chinensis,* which has a (possibly functional) copy of *ADHa*. In the previous chapter, we suggested that *Silene latifolia* represented an intermediate phenotype between Tyr- and Phe-dominant metabolism. *Simmondsia chinensis*, as a species with *ADHa* but not its most well-known consequence, betalain, may represent a similar occurrence where Tyr metabolism has not developed to the extent it has in betalain lineages, resulting in little overlap with betalain lineages or *ADHa*-lacking anthocyanin lineages. Alternatively, *Simmondsia* is the only desert species sampled, and may have evolved different strategies for high light stress than the other

three species by virtue of living in a different biome that typically has intense sunlight. The other three species are either partially domesticated or tend to live in disturbed areas around humans.

Without identifying the metabolites most impacting the OPLS-DA model, it is impossible to narrow these possibilities down. Tandem mass spectrometry and molecular networking is currently underway to further explore the metabolite diversity of these four species.

REFERENCES

Afendi, F.M., T. Okada, M. Yamazaki, A. Hirai-Morita, Y. Nakamura, K. Nakamura, S. Ikeda, et al. 2012. KNApSAck Family Databases: Integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53: e1.

Atanosov, A.G., S.B. Zotchev, V.M. Dirsch, the International Natural Product Sciences Taskforce, and C.T. Supuran. 2021. Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery* 20: 200-216.

Barros, J., J.C., Serrani-Yarce, F. Chen, D. Baxter, B.J Venables, amd R.A. Dixon. 2016. Role of bifunctional ammonia lyase in grass cell wall biosynthesis. *Nature Plants* 2: 16050.

Brereton, R.G. and G.R. Lloyd. 2014. Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics* 28: 213-225.

Busta, L., D. Hall, B. Johnson, M. Schaut, C.M. Hanson, A. Gupta, M. Gundrum, Y. Wang, H. Maeda. 2024. Mapping of specialized metabolite terms onto a plant phylogeny using text mining and large language models. *Plant J.*

Cho, M.-H., O.R.A Corea, H. Yang, D.L. Bedgar, D.D. Laskar, A.M. Anterola, F.A. Moog-Anterola, et al. 2007. Phenylalanine biosynthesis in *Arabidopsis thaliana*: Identification and characterization of arogenate dehydratases. *Journal of Biological Chemistry* 282: 30827-30835.

Clement, J.S., and T.J. Mabry. 1996. Pigment evolution in the Caryophyllales: a systematic overview. *Bot. Acta* 109: 360-367.

Gill, M. and W. Steglich. 1987. Pigments of fungi (Macromycetes). In *Fortschritte der Chemie*

Organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products, Vol. 51,W. Herz, H. Grisebach, G.W. Kirby, and C. Tamm, eds. (New York: Springer), pp. 1–297.

Gilman, I.S., J.J. Moreno-Villena, Z.R. Lewis, E.W. Goolsby, E.J. Edwards 2022. Gene co-expression reveals the modularity and integration of C_4 and CAM in *Portulaca. Plant Physiology* 189: 735-753.

Holeski, L.M., K. Keefover-Ring, J.M. Sobel, N.J. Kooyers. 2021. Evolutionary history and ecology shape the diversity and abundance of phytochemical arsenals across monkeyflowers. *Journal of Evolutionary Biology* 34: 571-583.

Huang, R., A.J. O'Donnell, J.J. Barboline, and T.J. Barkman 2016. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *PNAS* 113: 10613-10618.

Kim, H.-J., K.-J. Park, J.-H. Lim. 2011. Metabolomic analysis of phenolic compounds in buckwheat (*Fagopyrum esculentum*) sprouts treated with methyl jasmonate. *J. Agric. Food Chem.* 59: 5707-5713.

Landis, M., M. Tattini, and K. S. Gould. 2015. Multiple functional roles of anthocyanins in plant–environment interactions. *Environmental and Experimental Botany* 119: 4–17.

Lichman, B.R., G.T. Godden, J.P. Hamilton, L. Palmer, M.O. Kamileen, D. Zhao, B. Vaillancourt, et al. 2020. The evolutionary origins of the cat attractant nepetalactone in catnip. *Science Advances* 6: eaba 0721.

Lingwan, M., A.A. Pradhan, A.K Kushwaha, M.A. Dar, L. Bhagavatula, and S. Datta. 2023. Photoprotective role of plant secondary metabolites: Biosynthesis, photoregulation, and prospects of metabolic engineering for enhanced protection under excessive light. *Environmental and Experimental Botany* 209: 105300.

Liu, Q., L. Luo, L. Zheng. 2018. Lignins: Biosynthesis and biological functions in plants. *International Journal of Molecular Sciences* 19: 335.

Lopez-Nieves, S., Y. Yang, A. Timoneda, M. Wang, T. Feng, S. A. Smith, S. F. Brockington, and H. A. Maeda. 2018. Relaxation of tyrosine pathway regulation underlies the evolution of betalain pigmentation in Caryophyllales. *New Phytologist* 217: 896–908.

Jeon, J., J.K. Kim, Q. Wu, S.U. Park. 2018. Effects of cold stress on transcripts and metabolites in tartary buckwheat (*Fagopyrum tataricum*). *Environmental and Experimental Botany* 155: 488-496.

Martino, E., G. Casamassima, S. Castiglione, E. Cellupica, S. Pantalone, F. Papagni, M. Rui, A.M. Siciliano, S. Collina. 2018. Vinca alkaloids and analogues as anti-cancer agents: Looking back, peering ahead. *Bioorganic & Medicinal Chemistry Letters* 28: 2816-2826.

Meng, X., Z. Liang, X. Dai, Y. Zhang, S. Mahboub, D. W. Ngu. 2021. Predicting transcriptional responses to cold stress across plant species. *PNAS* 118: e2026330118.

Milech, C., P.A. Auler, M.N. do Amaral, S.R. Lucho, J.S. dos Santos, V.J,M. Furlan, V.J. Bianchi, E.J.B. Braga. 2023. Biosynthesis of betalain elicited by methyl jasmonate in two species of *Alternanthera* genus: Antagonistic regulations result in increase of pigments. *Applied Biochemistry and Biotechnology* 195: 4965-4982.

Moghe, G.D., and L.H. Kruse. 2018. The study of plant specialized metabolism: Challenges and prospects in the genomics era. *American Journal of Botany* 105: 959-962.

Moroldo, M., N. Blanchet, H. Duruflé, S. Bernillon, T. Berton, O. Fernandez, Y. Gibon, A. Moing, N.B. Langlade. 2024. Genetic control of abiotic stress-related specialized metabolites in sunflower. *BMC Genomics* 25: 199.

Pluskal, T., S. Castillo, A. Villar-Briones, M. Orešič. 2010. MZMine2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11: 395.

Rodman, J.E., P.S. Soltis, D.E. Soltis, K.J. Sytsma, and K.G. Karol. 1998. Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid phylogenies. *American Journal of Botany* 85: 997-1006.

Saccenti, E. and M.E. Timmerman. 2016. Approaches to sample size determination for multivariate data: application to PCA and PLS-DA of omics data. *Journal of Proteome Research* 15: 2379-2393.

Sharma, A., B. Shahzad, A. Rehman, R. Bhardwaj, M. Landi, B. Zheng. 2019. Response of Phenylpropanoid pathway and the role of polyphenols in plants under abiotic stress. *Molecules* 24: 2452.

Tiedge, K., X. Li, A.T. Merrill, D. Davisson, Y. Chen, P. Yu, D.J. Tantillo, R.L. Last, P. Zerbe. 2022. Comparative transcriptomics and metabolomics reveal specialized metabolite drought stress responses in switchgrass. *New Phytologist* 236: 1393-1408.

Tu, Y. 2011. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nature Medicine* 17: 1217-1220.

Wani, M.C., S.B. Horwitz. 2014. Nature as a remarkable chemist: a personal story of the discovery and development of Taxol. *Anti-Cancer Drugs* 25: 482-487.

Wasternack, C., Hause, B. 2013. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in Annals of Botany. *Annals of Botany* 111: 1021-1058.

Watkins, J.L., Q. Li, S. Yeaman, P.J. Facchini. 2023. Elucidation of the mescaline biosynthetic pathway in peyote (*Lophophora williamsii*). *The Plant Journal* 116: 635-649.

Weng, J-K., Lynch, J.H., Matos, J.O., Dudareva, N. 2021. Adaptive mechanisms of plant specialized metabolism connecting chemistry to function. *Nature Chemical Biology* 17: 1037-1045.

Wu, M., T.R. Northen, Y. Ding. 2023. Stressing the importance of plant specialized metabolites: omics-based approaches for discovering specialized metabolism in plant stress responses. *Frontiers in Plant Science* 14: 1272363.

Youssef, D., R. El-Bakatoushi, A. Elframawy, L. El-Sadek, and G. El Badan. 2023. Molecular phylogenetic study of flavonoids in medicinal plants: a case study of family Apiaceae. *Journal of Plant Research* 136: 305–322.

Zhang, J., J. Ge, B. Dayananda, J. Li. 2022. Effect of light intensities on the photosynthesis, growth, and physiological performances of two maple species. *Frontiers in Plant Science* 13: 999026.

Zhao, H., Y. Yang, S. Wang, X. Yang, K. Zhou, C. Xu, X. Zhang, et al. 2023. NPASS database update 2023. Quantitative natural product activity and species source database for biomedical research. *Nucleic Acids Research* 51: D621-D628.

CONCLUSION

In this dissertation, I have examined medicinal plants and specialized metabolism, with the Hmong community and the order Caryophyllales as study systems. I found that investigating specialized metabolism and medicinal plants at multiple scales can yield insight into their selection by humans and metabolic evolution.

In Chapter 1, I used DNA barcoding methods to identify plants used by the Hmong community, especially the Lee family, in Saint Paul, Minnesota for post-partum chicken soup. Incorporating sequence data and morphology allowed the identification of several herbs that were not identified in previous studies of the Hmong pharmacopeia. Additionally, the limits of such methods were made clear by the inability to identify a few species due to poor sequence quality or scattered taxonomic coverage in sequencing databases. Compared to other studies of Hmong herbs and other members of the community, I saw cultural variation and possible shifts in medicinal usage due to cultural drift and the Hmong diaspora. Thanks in part to this diaspora, broadening this study to more families and Hmong communities is an excellent system for understanding factors driving medicinal plant selection within a culture. Several of the medicinal species used by the Hmong also belonged to Caryophyllales.

In Chapter 2, I conducted a literature search on the medicinal use of selected Caryophyllales families across the globe and used phylogenetic comparative methods to determine that medicinal usage is not associated with Phe- or Tyr-dominant metabolism type. Instead, in Caryophyllales, the “apparency” or likelihood of interacting with humans was a better predictor of medicinal use across all categories of usage. Interestingly, the previous chapter’s discussion on cultural variation lends support to this global cross-cultural pattern, as the Hmong community incorporated introduced plants such as *Talinum paniculata* into their pharmacopeia

or substituted locally available plants in St. Paul that resembled ones used previously in Laos (i.e. *Iresine herbstii*). Assuming similar behavior holds in other cultures, it is easy to see how clades of closely related, widespread plants become preferentially used across cultures and medicinal categories.

Chapter 3 moved from a global scale and biodiversity methods, to the much smaller scale of two species and multi-omics methods. Using multiple treatments over a 48-hour time course, gene expression and semi-targeted metabolomics revealed that the interaction between the Tyr and Phe specialized metabolism pathways is more complex than previously thought, at least in our focal species. Rather than Tyr-derived metabolism functionally replacing the Phe-derived pathway in sea beet, the two pathways were most responsive to different treatments, suggesting that they evolved to mitigate different stressors. Furthermore, *Silene latifolia* does not appear to be a true reversal to Phe-dominant metabolism, instead maintaining remnants of both Phe and Tyr-enriched metabolism, although it remains to be seen whether this is unique to *Silene* or true across Caryophyllaceae.

Lastly, in Chapter 4, I used an untargeted metabolomics analysis to explore specialized metabolite diversity and response to high light and methyl jasmonate in four species (*Fagopyrum tataricum, Simmondsia chinensis, Dianthus barbatus,* and *Portulaca amilis*). While none of the four species significantly responded to methyl jasmonate, all four species induced different metabolites in response to high light stress, with very little overlap in the metabolites that responded. This indicates that there is considerable diversity in the specialized metabolic response to stress across Caryophyllales. Interestingly, overall metabolomics of the two species lacking *ADHα, Dianthus* and *Fagopyrum* were most similar, suggesting that the interactions between Phe and Tyr metabolism were driving some of the observed differences between

species. Taken together with the previous chapter, where *Silene latifolia* was shown to have an intermediate phenotype between the expected Phe-dominant and Tyr-dominant, this chapter suggests that, with broader taxonomic sampling across Caryophyllales, we may see different lineages balance Tyr and Phe metabolism in different ways.

The work described in this dissertation has contributed to our understanding of medicinal plant selection and specialized metabolism in Caryophyllales and generated considerable data for future investigation. Further work with the wider Hmong community can further disentangle the cultural factors contributing to the overall medicinal use patterns in Chapter 2. The multi-omics data generated in Chapter 3 is still a rich source for further investigation into Tyr and Phe metabolism, as well as other questions in these species. Discovery of genes correlated with the expression of the known genes explored in that chapter will widen our understanding of the diversity and regulation of these pathways in Caryophyllales, and yield insight into the production of bioactive specialized metabolites. Finally, greater taxonomic sampling and incorporating feature-based metabolomic networking techniques to expand the work outlined in Chapter 4 will better characterize the metabolomic diversity in the order and help inform our understanding of the relationship between gene expression and specialized metabolism.