

*PHYSCOMITRIUM PATENS*: APPLICATIONS IN SYNTHETIC BIOLOGY AND THE  
CURATION OF DITERPENOID LIBRARIES

By

Davis T. Mathieu

A DISSERTATION

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

Genetics & Genome Sciences — Doctor of Philosophy

2024

## ABSTRACT

The model moss species, *Physcomitrium patens* (*P. patens*), provides a unique system for investigating plant development, evolution, and physiology and also serves as an excellent chassis for synthetic biology. Because of the unique phylogeny of *P. patens* as a bryophyte, and sister to the vascular plants, allows opportunity for understanding shared traits among embryophytic life, early terrestrialization of plants on land 500 million years ago, and the divergences/convergences of plant traits since then. The high prevalence and global distribution of plants and fungi today can in-part be attributed to their long-standing relationship which predates early terrestrialization, and who's early collaboration likely reduced the initial barriers for both kingdoms to thrive on land. Here, the interaction between *P. patens* and fungi in the Mortierellaceae family are cocultured together to characterize their physiological and transcriptional responses. These analyses are used to explore possible long-standing interactions between plants and fungi, identify essential traits in plant-fungal communication, and provide foundational exploration into coculturing these systems for metabolite production. *P. patens* is an effective system for the production of heterologous metabolites, particularly diterpenes, because of its relatively low chemical diversity, many developed synthetic biological tools, and similar machinery/compartments to vascular plants. The large pool of diterpene chemodiversity and bioactivity known today affords these compounds with high humanitarian and economic value, making it an excellent metabolite to develop for expression in *P. patens*. Work presented in this dissertation focuses on evaluating the effectiveness of a coculture system with *P. patens* and Mortierellaceae, explores long-standing relationships among plants and fungi, provides schematic and initial testing of novel synthetic biological tools for improving *P. patens* as a chassis, and evaluates complexities among the total diterpene landscape to-date.

## TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES .....	vi
LIST OF ABBREVIATIONS.....	viii
<b>CHAPTER 1 <i>Physcomitrium patens</i>: A Chassis for Diterpene Synthesis and the Exploration of Fungal Symbiosis for Improved Growth and Extraction.....</b>	<b>1</b>
<i>Physcomitrium patens</i> as a Model Species and Chassis for Heterologous Expression: .....	2
Terpene Utility and Synthesis: .....	4
Native terpene metabolic pathways in <i>P. patens</i> :.....	7
Plant-Fungal Symbiosis:.....	8
Mortierellaceae characteristics, utility, and evolutionary relevance:.....	10
Work presented in this dissertation:.....	11
REFERENCES.....	14
<b>CHAPTER 2 Multilevel analysis between <i>Physcomitrium patens</i> and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi .....</b>	<b>27</b>
Abstract.....	28
Keywords.....	29
Significance Statement.....	29
Introduction.....	29
Methods & Materials .....	33
Results & Discussion .....	44
Conclusion.....	66
Data availability.....	67
Conflict of Interest .....	68
Acknowledgements.....	68
REFERENCES.....	70
<b>CHAPTER 3 Rule-Based Deconstruction and Reconstruction of the Diterpene Library: A Simulation of Synthesis and Unravelling of Compound Structural Diversity .....</b>	<b>85</b>
Abstract.....	86
Keywords.....	87
Significance Statement.....	87
Introduction.....	87
Methods & Materials .....	91
Results & Discussion .....	99
Conclusion.....	116
Data availability.....	118
Conflict of Interest .....	119
Acknowledgements.....	119
REFERENCES.....	121

<b>CHAPTER 4 Long Terminal Repeat Retrotransposon Targeted Transformation and Development of Promoter Reporter System in <i>Physcomitrium patens</i> for Sequential Targeting of Diterpene Module</b> .....	135
<b>Abstract</b> .....	136
<b>Keywords</b> .....	137
<b>Introduction</b> .....	137
<b>Materials &amp; Methods</b> .....	141
<b>Results</b> .....	147
<b>Conclusion &amp; Future Directions</b> .....	151
<b>REFERENCES</b> .....	152
 <b>APPENDIX: <i>CIRRICULUM VITAE</i></b> .....	 160

## LIST OF TABLES

<b>Table 2.1:</b> Select ontology and gene representatives from differential gene expression among <i>P. patens</i> and Mortierellaceae cocultures. ....	58
<b>Table 4.1:</b> Candidate genes for promoter cloning with <i>P. patens</i> reporter gene construct.....	151

## LIST OF FIGURES

<b>Figure 2.1:</b> Daily quantified growth of <i>P. patens</i> over 65 days compared to samples inoculated with either <i>B. erionia</i> WT, <i>B. erionia</i> CU, <i>L. elongata</i> WT, or <i>L. elongata</i> CU.....	45
<b>Figure 2.2:</b> Representative and specific interaction of <i>P. patens</i> with <i>B. erionia</i> and <i>L. elongata</i> (White bar indicates 50 $\mu$ m for each picture).....	47
<b>Figure 2.3:</b> Principal Component Analysis (PCA) of <i>P. patens</i> mapped RNA-seq reads for the 15 RNA-sequencing libraries generated with DeSEQ2. The color of each point correlates to experimental treatment.....	49
<b>Figure 2.4:</b> Venn diagram of DEGs between <i>P. patens</i> control and <i>P. patens</i> co-cultures with <i>B. erionia</i> WT (dark blue), <i>B. erionia</i> CU (light blue), <i>L. elongata</i> WT (red), and <i>L. elongata</i> CU (pink).....	50
<b>Figure 2.5:</b> Principal component analysis (PCA) of transcripts per million mapped reads (TPM) of the 99 libraries from the ‘Gene Atlas Project’ [Perroud et al. 2018] and the 15 samples analyzed here.....	62
<b>Figure 2.6:</b> Heatmap of the 822 genes with mapped reads to the dataset were presented here and absent among all ‘Gene Atlas Project’ samples.....	63
<b>Figure 2.7:</b> Mortierella coculture DEGs ( $P_{adj}<0.01$ , DESeq2) comparisons across <i>C. reinhardtii</i> , <i>A. thaliana</i> , and <i>P. patens</i> .....	66
<b>Figure 3.1:</b> Principal component analysis of DNP diterpene skeleton structures based on RDKit bit vector comparison scores.....	101
<b>Figure 3.2:</b> Summary of carbocation cyclization (TPS enzyme) reactions modeled at a global/theoretical level, filtered to identified structures, and examples of local synthesis.....	106
<b>Figure 3.3:</b> Summary of carbocation quenching patterns and post cyclization decoration for each of the top 20 most common diterpene skeleton classes in the TeroKit database.....	1097
<b>Figure 3.4:</b> Visual of atom and bond variation among the top 20 most common diterpene skeletons and identified variation related to diterpene origin, carbon connection(s), and carbon neighbor(s).....	112
<b>Figure 3.5:</b> Heatmap of top 50 most common diterpene skeletons and their abundance in plants and algae.....	115
<b>Figure 4.1:</b> Homologous DNA constructs for <i>P. patens</i> transformation.....	148

**Figure 4.2:** *Physcomitrium patens* transcript per million abundance of select genes grown in isolation and in coculture with either *B. erionia* or *L. elongata* ..... 149

**Figure 4.3:** Identified transcription factor binding motifs within the 5' UTR and 5kbp upstream of fungal responsive gene candidates ..... 150

## LIST OF ABBREVIATIONS

PEG	–Polyethylene Glycol
DNA	–Deoxynucleic Acid
SNP	–Single Nucleotide Polymorphism
BAC	– Bacterial Artificial Chromosome
Mbp	– Mega Base Pairs
LTR-RT	– Long Terminal Repeat Retrotransposon
BCE	– Before Common Era
DMAPP	– Dimethylallyl Diphosphate
IDP	– Isopentenyl Diphosphate
GGDP	– Geranylgeranyl Diphosphate
DNP	– Dictionary of Natural Products
TPS	– Terpene Synthase
diTPS	– Diterpene Synthase
OGD	–2-Oxoglutarate-Dependent Dioxygenase
UGT	– Uridine diphosphate glycosyltransferases
CPS/KS	– Copalyl Diphosphate/Kaurene Synthase
UV	– Ultraviolet
CRISPR/	– Clustered Regularly Interspaced Short Palindromic Repeats
Cas9	– CRISPR-Associated Protein 9
MRE	– Mollicutes-Related Endobacteria
BRE	– Burkholderia-Related Endobacteria
RNA	– Ribonucleic Acid

CERK1	– Chitin Elicitor Receptor Kinase 1
VAPYRIN	– Vesicle-Associated Membrane Protein (VAMP)-Associated Protein
AMF	– Arbuscular Mycorrhizal Fungi
WT	– Wild Type
CU	– Cured
MYA	– Million Years Ago
BCD	– Solution B, C, and D Media
MgSO <sub>4</sub>	– Magnesium Sulfate
KH <sub>2</sub> PO <sub>4</sub>	– Monopotassium Dihydrogenphosphate
KNO <sub>3</sub>	– Potassium Nitrate
FeSO <sub>4</sub> ·7H <sub>2</sub> O	– Ferrous Sulfate Heptahydrate
H <sub>2</sub> O	– Water
FAA	– Formalin-aceto-alcohol
EST	– Eastern Standard Time
LED	– Light-Emitting Diode
MSU HPCC	– Michigan State University High-Performance Computer Cluster
QC	– Quality Control
DEG	– Differentially Expressed Genes
P <sub>adj</sub>	– Adjusted P-value
RPKM	– Reads per Kilobase per Million Mapped Reads
TPM	– Transcripts per Million
PpNH <sub>4</sub>	– <i>Physcomitrium patens</i> Ammonium solution

PCA	– Principal Component Analysis
NCBI	– National Center for Biotechnology Information
PCR	– Polymerase Chain Reaction
CDS	– Coding Sequence
TAIR	– The Arabidopsis Information Resource
GO	– Gene Ontology
MLD-Kinase	– Myosin Light-Chain Kinase
CDPK	– Calcium-Dependent Protein Kinase
GRAS	– Gibberellic Acid Insensitive (GAI): Repressor of GAI (RGA): Scarecrow (SCR)
WRKY	– WRKY Binding Transcription Factor
TIFY	– TIFY Binding Transcription Factor
HEX	– Haematopoietically Expressed Homeobox Transcription Factor
HMG-CoA	– Hydroxymethylglutaryl-CoA Reductase
MVA	– Mevalonate
MYB-like	– Myeloblastosis-Like Transcription Factor
GDP	– Guanosine Diphosphate
ROS	– Reactive Oxygen Species
BIM1	– Binding to Microtubules protein
DOF	– DNA-Binding With One Finger Transcription Factor
ATP	– Adenosine Triphosphate
GLK1	– GOLDEN2-LIKE Transcription Factor
CaCO <sub>3</sub>	– Calcium Carbonate

MgCO <sub>3</sub>	– Magnesium Carbonate
siRNA	– Small Interfering Ribonucleic Acid
SMILES	– Simplified Molecular Input Line Entry System
SMARTS	– SMILES Arbitrary Target Specification
InChi	– IUPAC International Chemical Identifier
IQV	– Index of Qualitative Variation
H	– Hydrogen
C	– Carbon
CYP701	– Cytochrome P450; 701 family
YFP	– Yellow Fluorescent Protein
TE	– Transposable Elements
NPT-II	– Neomycin Phosphotransferase II
LP4-2A	– Linker Protein
OCS-T	– octopine synthase gene terminator
PNZ	– promoter module
Cfr9I	– <i>Citrobacter freundii</i> 9I endonuclease
AT	– Adenine/Thymine
GC	– Guanine/Cytosine
BLAST	– Basic Local Alignment Search Tool
CTAB	– Cetyltrimethylammonium bromide
EDTA	– Ethylenediaminetetraacetic Acid
TE buffer	– Tris-EDTA buffer

- gDNA – Genomic Deoxynucleic Acid
- IDT – Integrated DNA Technologies
- pBK3 – *Physcomitrium patens* CPS/KS Knockout
- GC/MS – Gas Chromatography/Mass Spectroscopy
- T<sub>m</sub> – Melting Temperature

## **CHAPTER 1**

### ***Physcomitrium patens*: A Chassis for Diterpene Synthesis and the Exploration of Fungal Symbiosis for Improved Growth and Extraction**

### ***Physcomitrium patens* as a Model Species and Chassis for Heterologous Expression:**

Native to North America, Europe, and Eastern Asia, *Physcomitrium patens* (*P. patens*; formerly *Physcomitrella patens*) has been a valuable tool aiding in plant biology research since its collection in 1962 [Engel 1968, Rensing et al. 2020]. The moss found initial utility in developmental plant genetics and plant hormone biology due to its morphological simplicity and phylogenetic relationship to vascular plants. The early production of multiple panels of mutant lines further accelerated its relevance within these fields [Engel, 1968, Ashton and Cove, 1977, Ashton et al., 1979, Abel et al., 1989]. The later development of a polyethylene-glycol (PEG)-mediated transformation in *P. patens* allowed for homologous recombination for the insertion of DNA into the genome, reigniting its utility within the study of plant biology and synthetic biology [Kammerer and Cove, 1996, Schaefer and Zrýd, 1997, Strepp et al., 1998]. Contrary to the vascular plants, *P. patens* and other bryophytes uniquely dominate the gametophytic, haploid life cycle. This dominant haploid cycle likely served as an important characteristic for the high transformation efficiency in *P. patens*, since only one gene copy must be inserted/knocked out to exhibit complete dominance. Capacity for transformation marked *P. patens* as one of the first multicellular organism to have transformation efficiency comparable to yeast (*Saccharomyces cerevisiae*) [Schaefer and Zrýd, 1997, Schaefer, 2001 Schaefer and Zrýd, 2001]. Specific examples of this included *P. patens* demonstrating capacity to produce the diterpene taxadiene [Anterola et al. 2009, Bach et al. 2014], and sesquiterpenes: patchoulol, and  $\alpha/\beta$ -santalene [Zhan et al. 2014] and artemisinin [Khairul Ikram et al. 2017]. Development of CRISPR/Cas9 technologies continued to expand the range of *P. patens* utility for heterologous expression, allowing for greater ease of multi-gene targeting and higher efficiency for gene knockout/insertion [Collonnier et al. 2017]. Continued improvements to transformation protocols

in *P. patens* have maintained the species as an excellent chassis for heterologous expression of biochemical pathways. Using *P. patens* in this way will be further explored in Chapter 4.

When considering a non-native chassis for heterologous expression, *P. patens* provides many benefits over both single celled and multi-cellular alternatives. An autotrophic lifestyle provides advantages when compared to yeast and bacteria, as it requires fewer inputs to maintain in a lab setting. Also, because most known specialized metabolites are derived from land plants, *P. patens* retains cellular phytochemical machinery and compartments more closely related to these sources compared to other, more popular platforms like bacteria, yeast, and algae [Fang et al. 2019, Zeng et al. 2020, Chapter 3]. *Physcomitrium patens* has flexibility in growing conditions and can be cultured in liquid, soil, and agar media, exhibits a high degree of stress tolerance, and can undergo storage indefinitely with cryopreservation [Schulte and Reski 2004, Frank et al. 2005, Mathieu et al. 2024 (Chapter 2)].

The turn of the millennia coincided largely with the next-generation sequencing revolution. Because *P. patens* had an established foothold at that time in 2008 led to it becoming the fifth land plant to have its genome sequenced [Rensing et al. 2008] following *Arabidopsis thaliana* [The Arabidopsis Genome Initiative 2000], rice (*Oryza sativa*) [International Rice Genome Sequencing Project 2005], poplar (*Populus tricharpa*) [Tuskan et al. 2006] and grape (*Vitis vinifera*) [The French–Italian Public Consortium for Grapevine Genome Characterization 2007]. This event established *P. patens* as the flagship genome for the second largest phylum of land plants, the Bryophyta [Rensing et al. 2008, Michael and Jackson 2013, Rensing et al. 2020]. This genome sequence was highly prioritized at the time due to its unique phylogenetic placement in relation to vascular plants and the algae *Chlamydomonas reinhardtii* (sequenced in 2007) [Merchant et al. 2007]. *Physcomitrium patens*' physiological similarities to the earliest land

plants referenced by fossil records further strengthened it as a platform for studying plant evolution [Kenrick and Crane 1997, Renzaglia and David 2001, Merchant et al. 2007, Rensing et al. 2008]. While *P. patens* has also adapted to the changing environment since its divergence from other land plants, it still has some capacity to act as a proxy for ancient land plants due to its shared morphology to ancestral accounts of land plants and an overall lower SNP accumulation over time compared to other known species such as *A. thaliana* [Lang et al. 2018]. In 2018, the *P. patens* genome was updated from its BAC/Fosmid predecessor to a 27 pseudochromosome level assembly using a shotgun sequencing strategy in combination with Sanger reads and the incorporation of former assemblies to create the 462 Mbp genome [Rensing et al. 2008, Lang et al 2018]. Publishing this new assembly resulted in the identification of unique genomic architecture compared to seed plants. Unlike most land plants, *P. patens* has fairly homogenous regions of genic space and long terminal repeat retrotransposon (LTR-RT) distribution across all chromosomes, instead of high gene distribution concentrated on chromosome arms [Lang et al. 2018]. This homogenization of genic and transposon spaces is likely also responsible for the abnormally high expression of LTR-RTs [Lang et al. 2018]. Despite high LTR-RT activity, *P. patens* seems to face minimal adverse effects and still retains a relatively small genome. This unique phenomenon is explored for its utility in synthetic biology applications in Chapter 4, particularly in the context of heterologous expression of diterpene pathways.

### **Terpene Utility and Synthesis:**

Terpenoids currently make up the largest known class of natural products with over 180,000 compounds reported in the TeroKit database as of July 2023<sup>1</sup> [Zeng et al. 2020, Zeng et al.

---

<sup>1</sup> <http://terokit.qmclab.com/data.html>

2022]. In nature, terpenoids function in organism-to-environment communication, particularly for clades with life cycles that are predominantly sessile, like that of plants, fungi, and corals [Zeng et al. 2020]. Plants produce the majority of terpenoid diversity, making up 75% of reported compounds [Zhou and Percherski 2020, Zeng et al. 2020], and find biological utility in many ways, including extracellular signaling [Irmisch et al. 2014, Dutta et al. 2017, Zeng et al. 2020, Rosenkranz et al. 2021], bacterial curation for the rhizosphere microbiome [Bullington et al. 2018, Huang and Osbourn 2019, Su et al. 2023], pollinator attraction [Kortbeek et al. 2019], and herbivory defense [Kortbeek et al. 2019, Ninkuu et al. 2021]. Humanity also has a long parallel history with terpene-producing plants, most notably in the form of traditional medicines. Some of the earliest written accounts of effective herbal medicines occur in the Babylonian Empire (~4,000 BCE) [Luqman 2014], the Yin and Shang dynasties in China (~1,000 BCE) [Ma et al. 2021], and ancient Greece (~500 BCE) [Jaiswal and Williams 2017]. While throughout human history many cultures have independently found uses for terpenes, western science still relies on the medicinal traditions and word of mouth inheritance of this information to continue compound discovery and understanding. In the US alone, terpenoid production is a multibillion-dollar industry, largely notable in their uses as fragrances, flavors, pharmaceuticals, nutraceuticals, and pesticides [Degenhardt et al. 2003, González-Coloma et al. 2014, Hausch et al. 2015, Koul 2008, Lange et al. 2011, Schalk et al. 2011, Phillippe et al. 2014, Celedon & Bohlmann 2016, Kutiyana & Bornemann 2018, Nuutinen 2018, Tetali 2019, Wang et al. 2005, Wani et al. 1971, Wilson & Roberts 2011, Zerbe et al. 2012, Zerbe and Bohlmann 2014, Zhao et al. 2016, Tetali 2019, Smith et al. 2022]. Future collaborations with many cultural groups must focus on respecting tradition, ethical sourcing of materials, and properly compensating all parties involved. If done properly,

these collaborations hold promise for the elucidation of many compounds that will likely provide services, utilities, and economic growth worldwide [Leonti and Casu 2013, Marks et al. 2023]. Diterpenoids originate from the combination of one dimethylallyl diphosphate (DMADP) and three isopentenyl diphosphates (IDP) to ultimately form the main diterpene precursor geranylgeranyl diphosphate (GGDP). Work presented here explores the role of the 20C diterpenoids in the context of heterologous expression in *P. patens* (Chapter 4) and reported diterpene diversity of compounds from the Dictionary of Natural Products (DNP; version 26.2) and TeroKit databases (Chapter 3) [Zeng et al 2019, Zeng et al. 2020, Zeng et al. 2022]. Nearly all reported diterpenes are derived from the combination of DMADP and three units of IDP via head-to-tail synthesis to create geranylgeranyl diphosphate (GGDP) [Schmidt et al. 2005]. Generally, GGDP is further modified with ring formation through help of diterpene synthases (diTPSs) to form various terpene backbones [Karunanithi and Zerbe 2019, Johnson et al. 2019]. Diterpene synthases can generally be classified into two modes of synthesis; either by a Class II/Class I mechanism or just a Class I enzyme acting alone, determined by which enzymes are present and which order reactions are carried out to form and resolve carbocations via cyclization, rearrangement, and elimination reactions [Karunanithi and Zerbe 2019]. The Class II/Class I mechanisms uses two enzymes (with some exceptions like in the case of *P. patens* in which one enzyme carries out both mechanisms) and first forms a carbocation at the GGDP tail (Class II) leading to initial cyclization. This is followed by the removal of the diphosphate (Class I), which yields another carbocation and additional cyclization or be resolved with the quenching of water or hydride ion. Alternatively, Class I diTPSs act independently to form and resolve the carbocation cascade reactions in one step starting by removing the diphosphate group. After cyclization, additional enzymes such as Cytochrome P450s (P450s), 2OGDs, UDP dependent

glycosyl transferases (UGT), and amino transferases provide the majority of observed structural diversity, which is explored more in Chapter 3.

### **Native terpene metabolic pathways in *P. patens*:**

*Physcomitrium patens* has low chemical diversity compared to most land plants [Bach et al. 2014]. Evidence for this is particularly exemplified by the absence of P450s and UGTs, which are generally responsible for the decoration of specialized metabolites and correlate with chemical diversity [Hamberger and Bak 2013, Nelson and Werck-Reichhart 2011]. While *A. thaliana* and *O. sativa* have 246 and 343 P450s respectively, *P. patens* only has 71 [Hamberger and Bak 2013]. This lack of background metabolite expression in *P. patens* strengthens it as a platform for the production of metabolites of interest since it reduces the risk of nonspecific modifications to heterologous products as a result of these enzymes having an abnormally high degree of substrate promiscuity [Nelson and Werck-Reichhart 2011, Hamberger and Bak 2013, Bach et al. 2014]. With regards to diterpene biosynthesis, *P. patens* has only one known endogenous diterpene pathway, which produces *ent*-kaurenoic acid, a precursor for the phytohormone gibberellin in vascular plants [Richards et al. 2001, Sun and Gubler 2004, Genschik 2009, Miyazaki et al. 2011, Davière and Achard 2013]. In *P. patens*, *ent*-kaurene synthesis is carried out by a single, bi-functional Class II/Class I diTPS copalyl diphosphate/kaurene synthase (CPS/KS) enzyme, which produces *ent*-kaur-16-ene and the hydroxylated isomer, *ent*-16- $\alpha$ -hydroxy kaurene [Hayashi et al. 2010, Hoffmann et al. 2014, Zhan et al. 2014]. *Physcomitrium patens* synthesizes other terpenoid classes such as the carotenoid derived strigolactone [Hoffmann et al. 2014], but *ent*-kaurene is by far the major terpene produced.

Although native terpene diversity is low, *P. patens* is excellent at producing large quantities of *ent*-kaurene, reaching concentrations 0.37-fold that of chlorophyll a and b [Zhan et al. 2014]. Despite the importance placed on *ent*-kaurene production, it is surprisingly not essential for cell growth [Hayashi et al. 2010]. *CPS/KS* knockout lines have a reduced capacity for cellular elongation, with stalled differentiation of chloronema (non-leafy undifferentiated filamentous tissue) but overall see little to no reduction of biomass [Hayashi et al. 2010]. While the differentiation of chloronema to caulonemata (the filamentous tissue preceding reproductive maturity) is important to reach sexual maturity in nature, chloronema can be maintained and grown indefinitely in a laboratory setting [Miyazaki et al. 2014]. From a biochemical standpoint, the endogenous diTPS pathway present in *P. patens* provides a nearly perfect foundation for the heterologous production of diterpenes due to high production of the GGDP precursor intended for *ent*-kaurene synthesis, low risk of nonspecific modification from P450s and UGTs, and the *ent*-kaurene knockouts arresting the organism at the easily cultured and genetically stable chloronema phase.

### **Plant-Fungal Symbiosis:**

About 90% of terrestrial plants share a mutualistic relationship with fungal symbionts [Bonfante and Genre 2010]. This interaction is generally characterized by an exchange of nutrients wherein plants provide a steady source of carbon through lipid or sugar derivatives, and in turn, fungi provide useable nitrogen, phosphorous, micronutrients, and improved water retention [Bonfante & Genre 2010, Martin & Nehls 2009]. Plant-fungal cocultures are also noted for better mitigation of stresses such as oxidative, osmotic, heat, UV radiation, and rapid temperature flux [de Vries & Archibald 2018, Du et al. 2019, Fürst-Jansen et al. 2020, Jermy 2011, Kohler et al. 2015, Lutzoni et al. 2018]. These stressors would have also posed as strong preventative forces

to initial plant terrestrialization as well. The observed stresses fungi and plants help mitigate together, the expansivity of environments in which they interact, and the traits each kingdom has to communicate with one another has led to the hypothesis that fungal mutualism may have played a critical role in initial plant colonization of land [Du et al. 2019, Knack et al. 2015, Kohler et al. 2015, Hanke & Rensing 2010, Liepina 2012, Loron et al. 2019, Lutzoni et al. 2001, Lutzoni et al. 2018, Morris et al. 2018, Nelson et al. 2020, Russell & Bulman 2005]. This symbiosis is further exemplified by the origin of land plants being tied largely to embryophytes, a clade with many connections to fungal symbiosis, suggesting that the necessary traits may have predated their emergence on land [Zhong et al. 2015, Morris et al. 2018]. The fossil records further support this, where key plant-fungal structures for interaction are observed dating back 407 million years ago (MYA) [Strullu-Derrien et al. 2014]. This parallels the emergence/retention of gene homologies necessary to interact with mycorrhizal structures before the divergence between charophytic algae and embryophytes (600 MYA) [Karandashov et al. 2004, Wang et al. 2010, Delaux et al. 2015].

Ancestral reconstruction of plant-fungal interaction predicts that the earliest plant-fungal symbionts likely resembled a common ancestor of Mucoraceae, Mortierellaceae and Glomeraceae, which all dominate the landscape today [Feijen et al. 2018]. Chapter 2 explores the shared physiology of early land plants and fungi compared with *P. patens* and specifically Mortierellaceae species, with aims to identify shared communication patterns throughout fungal-embryophyte relations today. The potential for symbiosis is also explored due to observed mutualistic interactions with algae, *A. thaliana* and other modern embryophytes [Becker & Cubeta 2020, Du et al. 2019, Feijen et al. 2018, Rensing et al. 2008, VandePol et al. 2022, Zhang et al. 2021]. Fungi have many applications in plant synthetic biology, where they have previously

been used as inoculants for improved growth and resiliency [Du et al. 2019, Vandepol et al. 2022], for the flocculation of algae out of solution [Jo et al. 2023, Shitanaka et al. 2023, Zhang et al. 2024], and for the biosynthesis hydrophobic lipid droplets in the storage of hydrophobic metabolites [Kamisaka et al. 1999, Yu et al. 2017].

### **Mortierellaceae characteristics, utility, and evolutionary relevance:**

Mortierellaceae represents a clade of globally ubiquitous fungi with isolates spanning from both the Antarctic [Weinstein et al. 2000] and Arctic poles [Gams et al. 1972, Salt 1977]. This family of soil-inhabiting fungi are characterized by their capacity for chitinolytic decomposition [Jackson 1965, Schlegel and Zaborosch 1993], prolific growth, multinucleated haploid mycelia, and bidirectional cytoplasmic transport mechanisms [Uehling et al. 2017]. Mortierellaceae have been found to be saprotrophic and associated with plant roots [Bonito et al. 2016, Liao et al. 2019] and have been reported to provide benefits to plant growth in corn (*Zea mays*), tomato (*Solanum lycopersicum*), watermelon (*Citrullus lanatus*), *A. thaliana*, and *C. reinhardtii* to name a few [Weber and Tribe 2003, Dyal and Narine 2005, Ueling et al. 2017, Du et al. 2019, Zhang et al. 2020, Telagathoti et al. 2021, Vandepol et al. 2022]. Because Mortierellaceae can easily be cultured, have capacity for mutualism, and are noted for their prolific and abundant production of essential fatty acids (50-80% fungal dry weight) [Pillai et al. 1998, Jansa et al. 1999, Zhang et al. 2021, Chang et al. 2022], pursuits to advance Mortierellaceae utility for synthetic biology have been of growing interest.

In the context of terpenoid heterologous expression in *P. patens*, Mortierellaceae could provide a potential benefit as a symbiont to accelerate plant growth, biomass production, and improved nutrient turnover. The substantial production of lipids and lipid droplets also could provide excellent vessels for the storage and extraction of hydrophobic terpenoids. On the basis of

phylogeny and morphology, the family Mortierellaceae share many similarities to the bryophytes with respect to their global dispersal, shared evolutionary expansion, and traits resembling the earliest predicted land-dwelling fungi [Krings et al. 2013, Krings et al. 2014, Feijen et al. 2018, Wang et al. 2023]. Work presented in Chapter 2 and 4 explore the potential of using Mortierellaceae to serve as a fungal symbiont or aggregate for improved production of heterologous pathways in *P. patens*. Mortierellaceae are excellent candidates in these pursuits due to their benefits in producing high value and hydrophobic lipid droplets, their demonstrated capacity for plant-fungal mutualism, and phylogenetic similarities with *P. patens* that may have implications in longstanding interactions between kingdoms.

**Work presented in this dissertation:**

Work presented here aims to demonstrate *Physcomitrium patens*' utility as a chassis for heterologous expression, particularly for diterpene production, and to identify complex patterns hidden within diterpene databases. As a model species, *P. patens* has utility in many overlapping fields of plant biology. While the work here was investigated with the purpose of improving heterologous production of natural products in *P. patens*, it is unavoidable to ignore *P. patens* response in the context of plant physiology, evolution, and development. Chapter 2 explores the viability of moss-fungal symbiosis with downstream applications for the harmonious production of hydrophobic metabolites (in moss) in conjunction with production of lipid droplets for metabolite storage (in fungi). While still the overarching intent, Chapter 2 also explores the complex physiological interactions of *P. patens* and Mortierellaceae from both the cellular and organismal level, explores the symbiotic tendencies of bryophytes and Mortierellaceae, and compares many transcriptional datasets to elucidate *P. patens* exchange with its environment. Chapter 4 takes advantage of the unique genomic architecture of *P. patens* and reduced

specialized metabolite background for heterologous expression of these two synthetic biological tools [Hamberger and Bak 2013, Lang et al. 2018, Banarjee et al. 2019]. The first tool investigates loci effects on metabolite production through the insertion of heterologous diterpene pathways, targeted to multiple LTR-RT loci. While likely not a viable solution in other systems, the unique LTR-RT distribution in the *P. patens* genome and relatively high LTR-RT activity provides expression comparable or in some cases outcompeting the Pp108 locus, considered neutral for the lack of detrimental impact upon insertion of heterologous genetic material, and thus a potential area for improved metabolite expression [Schaefer & Zryd 1997, Lang et al. 2018, Banerjee et al. 2019]. The second tool aims to produce a reporter system for testing the conditional response of *P. patens* promoters. This reporter system aims to identify promoters with consistently high expression compared to the more commonly used Ubiquitin promoter from *Zea mays*, which regulates a gene 450 million years diverged from *P. patens* [Christensen & Quail 1996]. Additionally, this system has potential for testing conditional regulatory effects, where a promoter is only active when certain parameters are met like when in the presence of fungi or later in development. *Physcomitrium patens* provides many advantages as a biological chassis for producing diterpenes. It has low metabolite diversity but a high diterpene production pool and the wealth of biological information that comes from being a model species makes this system a strong platform for diterpene production. The synthetic tools investigated here explore new avenues to further strengthen *P. patens* in these pursuits of metabolite production. Due to time constraints and challenges with cloning, this project was not completed, however, the necessary genetic elements have been assembled including large DNA fragments for insertion, CRISPR/Cas 9 vectors for Pp108 and LTR-RT loci and cloned out *P. patens* promoters. Chapter

4 will explore the steps to assemble the necessary genetic elements, challenges faced, and future directions for these new *P. patens* technologies.

Work presented in Chapter 3, investigates diterpene synthesis from a different perspective. A computational approach is taken to examine the wealth of reported diterpene diversity among the DNP and TeroKit databases (>60,000). Plants are the most commonly reported producers of unique terpenes and here it is explored further as to where that diversity occurs in the context of diterpene synthesis mechanisms and phylogenetic distributions. Major diterpene compound classes are dissected to determine where reported variability is presented in specific molecule groups, mainly in the context of alternative cyclization and/or functional group decoration. Curation strategies are presented for the analyzed diterpene databases, with the intent of improving these repositories in the future and identifying hidden, complex patterns. The explanation and sharing of novel modular software also provides new perspectives and ways to investigate compounds at an individual and a macroscopic scale. The modularity in software design comes from the plug-and-play approach so further refining can be done as discovery leads to more compounds and more information.

## REFERENCES

- Abel, W. O., Knebel, W., Koop, H.-U., Marienfeld, J. R., Quader, H., Reski, R., Schnepf, E., & Spörlein, B. (1989). A cytokinin-sensitive mutant of the moss, *Physcomitrella patens*, defective in chloroplast division. *Protoplasma*, *152*(1), 1–13. <https://doi.org/10.1007/BF01354234>
- Achard, P., & Genschik, P. (2009). Releasing the brakes of plant growth: How GAs shutdown DELLA proteins. *Journal of Experimental Botany*, *60*(4), 1085–1092. <https://doi.org/10.1093/jxb/ern301>
- Ashton, N. W., & Cove, D. J. (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Molecular and General Genetics MGG*, *154*(1), 87–95. <https://doi.org/10.1007/BF00265581>
- Ashton, N. W., Grimsley, N. H., & Cove, D. J. (1979). Analysis of gametophytic development in the moss, *Physcomitrella patens*, using auxin and cytokinin resistant mutants. *Planta*, *144*(5), 427–435. <https://doi.org/10.1007/BF00380118>
- Anterola, A., Shanle, E., Perroud, P.-F., & Quatrano, R. (2009). Production of taxa-4(5),11(12)-diene by transgenic *Physcomitrella patens*. *Transgenic Research*, *18*(4), 655–660. <https://doi.org/10.1007/s11248-009-9252-5>
- Bach, S. S., King, B. C., Zhan, X., Simonsen, H. T., & Hamberger, B. (2014). Heterologous Stable Expression of Terpenoid Biosynthetic Genes Using the Moss *Physcomitrella patens*. In M. Rodríguez-Concepción (Ed.), *Plant Isoprenoids: Methods and Protocols* (pp. 257–271). Springer. [https://doi.org/10.1007/978-1-4939-0606-2\\_19](https://doi.org/10.1007/978-1-4939-0606-2_19)
- Banerjee, A., Arnesen, J. A., Moser, D., Motsa, B. B., Johnson, S. R., & Hamberger, B. (2019). Engineering modular diterpene biosynthetic pathways in *Physcomitrella patens*. *Planta*, *249*(1), 221–233. <https://doi.org/10.1007/s00425-018-3053-0>
- Becker, L. E., & Cubeta, M. A. (2020). Increased Flower Production of *Calibrachoa x hybrida* by the Soil Fungus *Mortierella elongata*. *Journal of Environmental Horticulture*, *38*(4), 114–119. <https://doi.org/10.24266/0738-2898-38.4.114>
- Bonfante, P., & Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications*, *1*(1), Article 1. <https://doi.org/10.1038/ncomms1046>
- Bonito, G., Hameed, K., Ventura, R., Krishnan, J., Schadt, C. W., & Vilgalys, R. (2016). Isolating a functionally relevant guild of fungi from the root microbiome of *Populus*. *Fungal Ecology*, *22*, 35–42. <https://doi.org/10.1016/j.funeco.2016.04.007>

- Bullington, L. S., Lekberg, Y., Sniezko, R., & Larkin, B. (2018). The influence of genetics, defensive chemistry and the fungal microbiome on disease outcome in whitebark pine trees. *Molecular Plant Pathology*, *19*(8), 1847–1858. <https://doi.org/10.1111/mpp.12663>
- Celedon, J. M., & Bohlmann, J. (2016). Genomics-Based Discovery of Plant Genes for Synthetic Biology of Terpenoid Fragrances: A Case Study in Sandalwood oil Biosynthesis. *Methods in Enzymology*, *576*, 47–67. <https://doi.org/10.1016/bs.mie.2016.03.008>
- Chang, L., Lu, H., Chen, H., Tang, X., Zhao, J., Zhang, H., Chen, Y. Q., & Chen, W. (2022). Lipid metabolism research in oleaginous fungus *Mortierella alpina*: Current progress and future prospects. *Biotechnology Advances*, *54*, 107794. <https://doi.org/10.1016/j.biotechadv.2021.107794>
- Christensen, A. H., & Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research*, *5*(3), 213–218. <https://doi.org/10.1007/BF01969712>
- Collonnier, C., Epert, A., Mara, K., Maclot, F., Guyon-Debast, A., Charlot, F., White, C., Schaefer, D. G., & Nogu , F. (2017). CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss *Physcomitrella patens*. *Plant Biotechnology Journal*, *15*(1), 122–131. <https://doi.org/10.1111/pbi.12596>
- Davi re, J.-M., & Achard, P. (2013). Gibberellin signaling in plants. *Development*, *140*(6), 1147–1151. <https://doi.org/10.1242/dev.087650>
- de Vries, J., & Archibald, J. M. (2018). Plant evolution: Landmarks on the path to terrestrial life. *New Phytologist*, *217*(4), 1428–1434. <https://doi.org/10.1111/nph.14975>
- Degenhardt, J., Gershenzon, J., Baldwin, I. T., & Kessler, A. (2003). Attracting friends to feast on foes: Engineering terpene emission to make crop plants more attractive to herbivore enemies. *Current Opinion in Biotechnology*, *14*(2), 169–176. [https://doi.org/10.1016/S0958-1669\(03\)00025-9](https://doi.org/10.1016/S0958-1669(03)00025-9)
- Delaux, P.-M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., ... An , J.-M. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(43), 13390–13395. <https://doi.org/10.1073/pnas.1515426112>
- Du, Z.-Y., Zienkiewicz, K., Vande Pol, N., Ostrom, N. E., Benning, C., & Bonito, G. M. (2019). Algal-fungal symbiosis leads to photosynthetic mycelium. *eLife*, *8*, e47815. <https://doi.org/10.7554/eLife.47815>

- Dutta, S., Mehrotra, R. C., Paul, S., Tiwari, R. P., Bhattacharya, S., Srivastava, G., Ralte, V. Z., & Zoramthara, C. (2017). Remarkable preservation of terpenoids and record of volatile signalling in plant-animal interactions from Miocene amber. *Scientific Reports*, 7(1), Article 1. <https://doi.org/10.1038/s41598-017-09385-w>
- Dyal, S. D., & Narine, S. S. (2005). Implications for the use of *Mortierella* fungi in the industrial production of essential fatty acids. *Food Research International*, 38(4), 445–467. <https://doi.org/10.1016/j.foodres.2004.11.002>
- Engel, P. P. (1968). The Induction of Biochemical and Morphological Mutants in the Moss *Physcomitrella patens*. *American Journal of Botany*, 55(4), 438–446. <https://doi.org/10.2307/2440573>
- Fang, C., Fernie, A. R., & Luo, J. (2019). Exploring the Diversity of Plant Metabolism. *Trends in Plant Science*, 24(1), 83–98. <https://doi.org/10.1016/j.tplants.2018.09.006>
- Feijen, F. A. A., Vos, R. A., Nuytinck, J., & Merckx, V. S. F. T. (2018). Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Scientific Reports*, 8(1), Article 1. <https://doi.org/10.1038/s41598-018-28920-x>
- Frank, W., Ratnadewi, D., & Reski, R. (2005). *Physcomitrella patens* is highly tolerant against drought, salt and osmotic stress. *Planta*, 220(3), 384–394. <https://doi.org/10.1007/s00425-004-1351-1>
- Fürst-Jansen, J. M. R., de Vries, S., & de Vries, J. (2020). Evo-physio: On stress responses and the earliest land plants. *Journal of Experimental Botany*, 71(11), 3254–3269. <https://doi.org/10.1093/jxb/eraa007>
- Gams, W., Chien, C.-Y., & Domsch, K. H. (1972). Zygosporangium formation by the heterothallic *Mortierella elongata* and a related homothallic species, *M. epigama* sp. nov. *Transactions of the British Mycological Society*, 58(1), 5-IN2. [https://doi.org/10.1016/S0007-1536\(72\)80065-2](https://doi.org/10.1016/S0007-1536(72)80065-2)
- González-Coloma, A., Guadaño, A., Tonn, C. E., & Sosa, M. E. (2005). Antifeedant/Insecticidal Terpenes from Asteraceae and Labiatae Species Native to Argentinean Semi-arid Lands. *Zeitschrift Für Naturforschung C*, 60(11–12), 855–861. <https://doi.org/10.1515/znc-2005-11-1207>
- Hamberger, B., & Bak, S. (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1612), 20120426. <https://doi.org/10.1098/rstb.2012.0426>
- Hanke, S., & Rensing, S. (2010). In vitro association of non-seed plant gametophytes with arbuscular mycorrhiza fungi. *Endocytobiosis and Cell Research*, 20, 95–101.

- Hausch, B. J., Lorjaroenphon, Y., & Cadwallader, K. R. (2015). Flavor chemistry of lemon-lime carbonated beverages. *Journal of Agricultural and Food Chemistry*, 63(1), 112–119. <https://doi.org/10.1021/jf504852z>
- Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A., Nakashima, T., Nakajima, M., Mander, L. N., Yamane, H., Hasebe, M., & Nozaki, H. (2010). Endogenous Diterpenes Derived from *ent*-Kaurene, a Common Gibberellin Precursor, Regulate Protonema Differentiation of the Moss *Physcomitrella patens*. *Plant Physiology*, 153(3), 1085–1097. <https://doi.org/10.1104/pp.110.157909>
- Hoffmann, B., Proust, H., Belcram, K., Labrune, C., Boyer, F.-D., Rameau, C., & Bonhomme, S. (2014). Strigolactones Inhibit Caulonema Elongation and Cell Division in the Moss *Physcomitrella patens*. *PLOS ONE*, 9(6), e99206. <https://doi.org/10.1371/journal.pone.0099206>
- Huang, A. C., & Osbourn, A. (2019). Plant terpenes that mediate below-ground interactions: Prospects for bioengineering terpenoids for plant protection. *Pest Management Science*, 75(9), 2368–2377. <https://doi.org/10.1002/ps.5410>
- International Rice Genome Sequencing Project, & Sasaki, T. (2005). The map-based sequence of the rice genome. *Nature*, 436(7052), 793–800. <https://doi.org/10.1038/nature03895>
- Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J., & Köllner, T. G. (2014). Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). *BMC Plant Biology*, 14(1), 270. <https://doi.org/10.1186/s12870-014-0270-y>
- Jackson, R. M. (1965). Studies of fungi in pasture soils. *New Zealand Journal of Agricultural Research*, 8(4), 878–888. <https://doi.org/10.1080/00288233.1965.10423722>
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguene, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., ... The French–Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161), 463–467. <https://doi.org/10.1038/nature06148>
- Jaiswal, Y. S., & Williams, L. L. (2017). A glimpse of Ayurveda – The forgotten history and principles of Indian traditional medicine. *Journal of Traditional and Complementary Medicine*, 7(1), 50–53. <https://doi.org/10.1016/j.jtcme.2016.02.002>
- Jansa, J., Gryndler, M., & Matucha, M. (1999). Comparison of the lipid profiles of arbuscular mycorrhizal (AM) fungi and soil saprophytic fungi. *Symbiosis*.
- Jermey, A. (2011). Soil fungi helped ancient plants to make land. *Nature Reviews Microbiology*, 9(1), Article 1. <https://doi.org/10.1038/nrmicro2494>

- Jo, C., Zhang, J., Tam, J. M., Church, G. M., Khalil, A. S., Segrè, D., & Tang, T.-C. (2023). Unlocking the magic in mycelium: Using synthetic biology to optimize filamentous fungi for biomanufacturing and sustainability. *Materials Today Bio*, *19*, 100560. <https://doi.org/10.1016/j.mtbio.2023.100560>
- Kamisaka, Y., Noda, N., Sakai, T., & Kawasaki, K. (1999). Lipid bodies and lipid body formation in an oleaginous fungus, *Mortierella ramanniana* var. *Angulispora*. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1438*(2), 185–198. [https://doi.org/10.1016/S1388-1981\(99\)00050-5](https://doi.org/10.1016/S1388-1981(99)00050-5)
- Kammerer, W., & Cove, D. J. (1996). Genetic analysis of the effects of re-transformation of transgenic lines of the moss *Physcomitrella patens*. *Molecular and General Genetics MGG*, *250*(3), 380–382. <https://doi.org/10.1007/BF02174397>
- Karandashov, V., Nagy, R., Wegmüller, S., Amrhein, N., & Bucher, M. (2004). Evolutionary conservation of a phosphate transporter in the arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 6285–6290. <https://doi.org/10.1073/pnas.0306074101>
- Karunanithi, P. S., & Zerbe, P. (2019). Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Frontiers in Plant Science*, *10*. <https://www.frontiersin.org/articles/10.3389/fpls.2019.01166>
- Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*, *389*(6646), Article 6646. <https://doi.org/10.1038/37918>
- Khairul Ikram, N. K. B., Beyraghdar Kashkooli, A., Peramuna, A. V., van der Krol, A. R., Bouwmeester, H., & Simonsen, H. T. (2017). Stable Production of the Antimalarial Drug Artemisinin in the Moss *Physcomitrella patens*. *Frontiers in Bioengineering and Biotechnology*, *5*. <https://www.frontiersin.org/articles/10.3389/fbioe.2017.00047>
- Kluger, R., & Eastman, R. (2018). Isoprenoid: Structural features of Isoprenoids. *Encyclopedia Britannica*. <https://www.britannica.com/science/isoprenoid>
- Knack, J. J., Wilcox, L. W., Delaux, P.-M., Ané, J.-M., Piotrowski, M. J., Cook, M. E., Graham, J. M., & Graham, L. E. (2015). Microbiomes of Streptophyte Algae and Bryophytes Suggest That a Functional Suite of Microbiota Fostered Plant Colonization of Land. *International Journal of Plant Sciences*, *176*(5), 405–420. <https://doi.org/10.1086/681161>
- Kohler, A., Kuo, A., Nagy, L. G., Morin, E., Barry, K. W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M. D., Doré, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., ... Martin, F. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics*, *47*(4), Article 4. <https://doi.org/10.1038/ng.3223>

- Kortbeek, R. W. J., van der Gragt, M., & Bleeker, P. M. (2019). Endogenous plant metabolites against insects. *European Journal of Plant Pathology*, 154(1), 67–90. <https://doi.org/10.1007/s10658-018-1540-6>
- Koul, O. (2008). Phytochemicals and Insect Control: An Antifeedant Approach. *Critical Reviews in Plant Sciences*, 27(1), 1–24. <https://doi.org/10.1080/07352680802053908>
- Krings, M., Taylor, T. N., & Dotzler, N. (2013). Fossil evidence of the zygomycetous fungi. *Persoonia - Molecular Phylogeny and Evolution of Fungi*, 30(1), 1–10. <https://doi.org/10.3767/003158513X664819>
- Krings, M., Taylor, T. N., Taylor, E. L., Kerp, H., & Dotzler, N. (2014). First record of a fungal “sporocarp” from the Lower Devonian Rhynie chert. *Palaeobiodiversity and Palaeoenvironments*, 94(2), 221–227. <https://doi.org/10.1007/s12549-013-0135-7>
- Kutyna, D. R., & Borneman, A. R. (2018). Heterologous Production of Flavour and Aroma Compounds in *Saccharomyces cerevisiae*. *Genes*, 9(7), Article 7. <https://doi.org/10.3390/genes9070326>
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R., Vives, C., Morata, J., Symeonidi, A., Hiss, M., Muchero, W., Kamisugi, Y., Saleh, O., Blanc, G., Decker, E. L., ... Rensing, S. A. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, 93(3), 515–533. <https://doi.org/10.1111/tpj.13801>
- Lange, B. M., Mahmoud, S. S., Wildung, M. R., Turner, G. W., Davis, E. M., Lange, I., Baker, R. C., Boydston, R. A., & Croteau, R. B. (2011). Improving peppermint essential oil yield and composition by metabolic engineering. *Proceedings of the National Academy of Sciences*, 108(41), 16944–16949. <https://doi.org/10.1073/pnas.1111558108>
- Lemmich, E. (1979). Peucelinendiol, a new acyclic diterpenoid from *Peucedanum oreoselinum*. *Phytochemistry*, 18(7), 1195–1197. [https://doi.org/10.1016/0031-9422\(79\)80133-8](https://doi.org/10.1016/0031-9422(79)80133-8)
- Leonti, M., & Casu, L. (2013). Traditional medicines and globalization: Current and future perspectives in ethnopharmacology. *Frontiers in Pharmacology*, 4, 92. <https://doi.org/10.3389/fphar.2013.00092>
- Liao, H.-L., Bonito, G., Rojas, J. A., Hameed, K., Wu, S., Schadt, C. W., Labbé, J., Tuskan, G. A., Martin, F., Grigoriev, I. V., & Vilgalys, R. (2019). Fungal Endophytes of *Populus trichocarpa* Alter Host Phenotype, Gene Expression, and Rhizobiome Composition. *Molecular Plant-Microbe Interactions®*, 32(7), 853–864. <https://doi.org/10.1094/MPMI-05-18-0133-R>
- Liepiņa, L. (2021). *Occurrence of fungal structures in bryophytes of the boreo-nemoral zone*. 6.

- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., & Duckett, J. G. (2007). Glomeromycotean associations in liverworts: A molecular, cellular, and taxonomic analysis. *American Journal of Botany*, *94*(11), 1756–1777. <https://doi.org/10.3732/ajb.94.11.1756>
- Loron, C. C., François, C., Rainbird, R. H., Turner, E. C., Borensztajn, S., & Javaux, E. J. (2019). Early fungi from the Proterozoic era in Arctic Canada. *Nature*, *570*(7760), Article 7760. <https://doi.org/10.1038/s41586-019-1217-0>
- Luqman, S. (2014). The Saga of Opium Poppy: Journey from Traditional Medicine to Modern Drugs and Nutraceuticals. *Acta Horticulturae*, *1036*, 91–100. <https://doi.org/10.17660/ActaHortic.2014.1036.9>
- Lutzoni, F., Nowak, M. D., Alfaro, M. E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A. E., Lewis, L. A., Swofford, D. L., Hibbett, D., Hilu, K., James, T. Y., Quandt, D., & Magallón, S. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications*, *9*(1), Article 1. <https://doi.org/10.1038/s41467-018-07849-9>
- Lutzoni, F., Pagel, M., & Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, *411*(6840), 937–940. <https://doi.org/10.1038/35082053>
- Ma, D., Wang, S., Shi, Y., Ni, S., Tang, M., & Xu, A. (2021). The development of traditional Chinese medicine. *Journal of Traditional Chinese Medical Sciences*, *8*, S1–S9. <https://doi.org/10.1016/j.jtcms.2021.11.002>
- Marks, R. A., Amézquita, E. J., Percival, S., Rougon-Cardoso, A., Chibici-Revneanu, C., Tebele, S. M., Farrant, J. M., Chitwood, D. H., & VanBuren, R. (2023). A critical analysis of plant science literature reveals ongoing inequities. *Proceedings of the National Academy of Sciences*, *120*(10), e2217564120. <https://doi.org/10.1073/pnas.2217564120>
- Martin, F., & Nehls, U. (2009). Harnessing ectomycorrhizal genomics for ecological insights. *Current Opinion in Plant Biology*, *12*(4), 508–515. <https://doi.org/10.1016/j.pbi.2009.05.007>
- Mathieu, D., Bryson, A. E., Hamberger, B., Singan, V., Keymanesh, K., Wang, M., Barry, K., Mondo, S., Pangilinan, J., Koriabine, M., Grigoriev, I. V., Bonito, G., & Hamberger, B. (2024). Multilevel analysis between *Physcomitrium patens* and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi. *The Plant Journal*. <https://doi.org/10.1111/tpj.16605>
- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., Salamov, A., Fritz-Laylin, L. K., Maréchal-Drouard, L., Marshall, W. F., Qu, L.-H., Nelson, D. R., Sanderfoot, A. A., Spalding, M. H., Kapitonov, V. V., Ren, Q., Ferris, P., Lindquist, E., ... Grossman, A. R. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science (New York, N.Y.)*, *318*(5848), 245–250. <https://doi.org/10.1126/science.1143609>

- Michael, T. P., & Jackson, S. (2013). The First 50 Plant Genomes. *The Plant Genome*, 6(2), plantgenome2013.03.0001in. <https://doi.org/10.3835/plantgenome2013.03.0001in>
- Miller, G. P., Bhat, W. W., Lanier, E. R., Johnson, S. R., Mathieu, D. T., & Hamberger, B. (2020). The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all-cis prenyl intermediate. *The Plant Journal*, 104(3), 693–705. <https://doi.org/10.1111/tpj.14957>
- Miyazaki, S., Katsumata, T., Natsume, M., & Kawaide, H. (2011). The CYP701B1 of *Physcomitrella patens* is an *ent*-kaurene oxidase that resists inhibition by uniconazole-P. *FEBS Letters*, 585(12), 1879–1883. <https://doi.org/10.1016/j.febslet.2011.04.057>
- Miyazaki, S., Toyoshima, H., Natsume, M., Nakajima, M., & Kawaide, H. (2014). Blue-light irradiation up-regulates the *ent*-kaurene synthase gene and affects the avoidance response of protonemal growth in *Physcomitrella patens*. *Planta*, 240(1), 117–124. <https://doi.org/10.1007/s00425-014-2068-4>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10), E2274–E2283. <https://doi.org/10.1073/pnas.1719588115>
- Nelsen, M. P., Lücking, R., Boyce, C. K., Lumbsch, H. T., & Ree, R. H. (2020). No support for the emergence of lichens prior to the evolution of vascular plants. *Geobiology*, 18(1), 3–13. <https://doi.org/10.1111/gbi.12369>
- Nelson, D., & Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. *The Plant Journal*, 66(1), 194–211. <https://doi.org/10.1111/j.1365-313X.2011.04529.x>
- Ninkuu, V., Zhang, L., Yan, J., Fu, Z., Yang, T., & Zeng, H. (2021). Biochemistry of Terpenes and Recent Advances in Plant Protection. *International Journal of Molecular Sciences*, 22(11), 5710. <https://doi.org/10.3390/ijms22115710>
- Nuutinen, T. (2018). Medicinal properties of terpenes found in *Cannabis sativa* and *Humulus lupulus*. *European Journal of Medicinal Chemistry*, 157, 198–228. <https://doi.org/10.1016/j.ejmech.2018.07.076>
- Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. (2018). Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes*, 9(2), 88. <https://doi.org/10.3390/genes9020088>
- Peters, R. J. (2010). Two rings in them all: The labdane-related diterpenoids. *Natural Product Reports*, 27(11), 1521. <https://doi.org/10.1039/c0np00019a>

- Philippe, R. N., De Mey, M., Anderson, J., & Ajikumar, P. K. (2014). Biotechnological production of natural zero-calorie sweeteners. *Current Opinion in Biotechnology*, 26, 155–161. <https://doi.org/10.1016/j.copbio.2014.01.004>
- Pillai, M. G., Certik, M., Nakahara, T., & Kamisaka, Y. (1998). Characterization of triacylglycerol biosynthesis in subcellular fractions of an oleaginous fungus, *Mortierella ramanniana* var. *Angulispora*. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1393(1), 128–136. [https://doi.org/10.1016/S0005-2760\(98\)00069-1](https://doi.org/10.1016/S0005-2760(98)00069-1)
- Rensing, S. A., Goffinet, B., Meyberg, R., Wu, S.-Z., & Bezanilla, M. (2020). The Moss *Physcomitrium* (*Physcomitrella*) *patens*: A Model Organism for Non-Seed Plants. *The Plant Cell*, 32(5), 1361–1376. <https://doi.org/10.1105/tpc.19.00828>
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., ... Boore, J. L. (2008). The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science*, 319(5859), 64–69.
- Renzaglia, K. S., & Garbary, D. J. (2001). Motile Gametes of Land Plants: Diversity, Development, and Evolution. *Critical Reviews in Plant Sciences*, 20(2), 107–213. <https://doi.org/10.1080/20013591099209>
- Richards, D. E., King, K. E., Ait-ali, T., & Harberd, N. P. (2001). HOW GIBBERELLIN REGULATES PLANT GROWTH AND DEVELOPMENT: A Molecular Genetic Analysis of Gibberellin Signaling. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52(1), 67–88. <https://doi.org/10.1146/annurev.arplant.52.1.67>
- Rosenkranz, M., Chen, Y., Zhu, P., & Vlot, A. C. (2021). Volatile terpenes—Mediators of plant-to-plant communication. *The Plant Journal: For Cell and Molecular Biology*, 108(3), 617–631. <https://doi.org/10.1111/tpj.15453>
- Russell, J., & Bulman, S. (2005). The liverwort *Marchantia foliacea* forms a specialized symbiosis with arbuscular mycorrhizal fungi in the genus *Glomus*. *New Phytologist*, 165(2), 567–579. <https://doi.org/10.1111/j.1469-8137.2004.01251.x>
- Salt, G. A. (1977). The Incidence of Root-Surface Fungi on Naturally Regenerated *Picea sitchensis* Seedlings in Southeast Alaska. *Forestry: An International Journal of Forest Research*, 50(2), 113–115. <https://doi.org/10.1093/forestry/50.2.113>
- Schaefer, D. (2001). Gene targeting in *Physcomitrella patens*. *Current Opinion in Plant Biology*, 4(2), 143–150. [https://doi.org/10.1016/S1369-5266\(00\)00150-3](https://doi.org/10.1016/S1369-5266(00)00150-3)
- Schaefer, D. G., & Zryd, J.-P. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *The Plant Journal*, 11(6), 1195–1206. <https://doi.org/10.1046/j.1365-313X.1997.11061195.x>

- Schlegel, H. G., & Zaborosch, C. (1993). *General Microbiology*. Cambridge University Press.
- Schmidt, A., Zeneli, G., Hietala, A. M., Fossdal, C. G., Krokene, P., Christiansen, E., & Gershenzon, J. (2005). Chapter One - Induced Chemical Defenses in Conifers: Biochemical and Molecular Approaches to Studying Their Function. In J. T. Romeo (Ed.), *Recent Advances in Phytochemistry* (Vol. 39, pp. 1–28). Elsevier.  
[https://doi.org/10.1016/S0079-9920\(05\)80002-4](https://doi.org/10.1016/S0079-9920(05)80002-4)
- Schulte, J., & Reski, R. (2004). High Throughput Cryopreservation of 140,000 *Physcomitrella patens* Mutants. *Plant Biology*, 6(2), 119–127. <https://doi.org/10.1055/s-2004-817796>
- Shitanaka, T., Higa, L., Bryson, A. E., Bertucci, C., Vande Pol, N., Lucker, B., Khanal, S. K., Bonito, G., & Du, Z.-Y. (2023). Flocculation of oleaginous green algae with *Mortierella alpina* fungi. *Bioresource Technology*, 385, 129391.  
<https://doi.org/10.1016/j.biortech.2023.129391>
- Smith, C. J., Vergara, D., Keegan, B., & Jikomes, N. (2022). The phytochemical diversity of commercial Cannabis in the United States. *PLOS ONE*, 17(5), e0267498.  
<https://doi.org/10.1371/journal.pone.0267498>
- Strepp, R., Scholz, S., Kruse, S., Speth, V., & Reski, R. (1998). Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proceedings of the National Academy of Sciences of the United States of America*, 95(8), 4368–4373.
- Strullu-Derrien, C., Kenrick, P., Pressel, S., Duckett, J. G., Rioult, J.-P., & Strullu, D.-G. (2014). Fungal associations in *Horneophyton ligneri* from the Rhynie Chert (c. 407 million year old) closely resemble those in extant lower land plants: Novel insights into ancestral plant-fungus symbioses. *The New Phytologist*, 203(3), 964–979.  
<https://doi.org/10.1111/nph.12805>
- Su, J., Wang, Y., Bai, M., Peng, T., Li, H., Xu, H.-J., Guo, G., Bai, H., Rong, N., Sahu, S. K., He, H., Liang, X., Jin, C., Liu, W., Strube, M. L., Gram, L., Li, Y., Wang, E., Liu, H., & Wu, H. (2023). Soil conditions and the plant microbiome boost the accumulation of monoterpenes in the fruit of *Citrus reticulata* ‘Chachi.’ *Microbiome*, 11(1), 61.  
<https://doi.org/10.1186/s40168-023-01504-2>
- Sun, T., & Gubler, F. (2004). Molecular Mechanism of Gibberellin Signaling in Plants. *Annual Review of Plant Biology*, 55(1), 197–223.  
<https://doi.org/10.1146/annurev.arplant.55.031903.141753>
- Telagathoti, A., Probst, M., & Peintner, U. (2021). Habitat, Snow-Cover and Soil pH, Affect the Distribution and Diversity of Mortierellaceae Species and Their Associations to Bacteria. *Frontiers in Microbiology*, 12, 669784. <https://doi.org/10.3389/fmicb.2021.669784>

- Tetali, S. D. (2019). Terpenes and isoprenoids: A wealth of compounds for global use. *Planta*, 249(1), 1–8. <https://doi.org/10.1007/s00425-018-3056-x>
- The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), 796–815. <https://doi.org/10.1038/35048692>
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., ... Rokhsar, D. (2006). The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313(5793), 1596–1604. <https://doi.org/10.1126/science.1128691>
- Uehling, J., Gryganskyi, A., Hameed, K., Tschaplinski, T., Misztal, P. K., Wu, S., Desirò, A., Vande Pol, N., Du, Z., Zienkiewicz, A., Zienkiewicz, K., Morin, E., Tisserant, E., Splivallo, R., Hainaut, M., Henrissat, B., Ohm, R., Kuo, A., Yan, J., ... Bonito, G. (2017). Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycoavidus cysteinexigens*. *Environmental Microbiology*, 19(8), 2964–2983. <https://doi.org/10.1111/1462-2920.13669>
- Vandepol, N., Liber, J., Yocca, A., Matlock, J., Edger, P., & Bonito, G. (2022). *Linnemannia elongata* (Mortierellaceae) stimulates *Arabidopsis thaliana* aerial growth and responses to auxin, ethylene, and reactive oxygen species. *PLOS ONE*, 17(4), e0261908. <https://doi.org/10.1371/journal.pone.0261908>
- Wang, B., Yeun, L. H., Xue, J.-Y., Liu, Y., Ané, J.-M., & Qiu, Y.-L. (2010). Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. *The New Phytologist*, 186(2), 514–525. <https://doi.org/10.1111/j.1469-8137.2009.03137.x>
- Wang, G., Tang, W., & Bidigare, R. R. (2005). Terpenoids As Therapeutic Drugs and Pharmaceutical Agents. In L. Zhang & A. L. Demain (Eds.), *Natural Products: Drug Discovery and Therapeutic Medicine* (pp. 197–227). Humana Press. [https://doi.org/10.1007/978-1-59259-976-9\\_9](https://doi.org/10.1007/978-1-59259-976-9_9)
- Wang, Y., Chang, Y., Ortañez, J., Peña, J. F., Carter-House, D., Reynolds, N. K., Smith, M. E., Benny, G., Mondo, S. J., Salamov, A., Lipzen, A., Pangilinan, J., Guo, J., LaButti, K., Andreopolous, W., Tritt, A., Keymanesh, K., Yan, M., Barry, K., ... Stajich, J. E. (2023). Divergent Evolution of Early Terrestrial Fungi Reveals the Evolution of Mucormycosis Pathogenicity Factors. *Genome Biology and Evolution*, 15(4), evad046. <https://doi.org/10.1093/gbe/evad046>
- Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93(9), 2325–2327. <https://doi.org/10.1021/ja00738a045>

- Weber, R. W. S., & Tribe, H. T. (2003). Oil as a substrate for *Mortierella* species. *Mycologist*, 17(3), 134–139. [https://doi.org/10.1017/S0269-915X\(03\)00203-9](https://doi.org/10.1017/S0269-915X(03)00203-9)
- Wefer, J., Simon, K., & Lindel, T. (2013). Cubitane: A rare diterpenoid skeleton. *Phytochemistry Reviews*, 12(1), 95–105. <https://doi.org/10.1007/s11101-012-9257-1>
- Weinstein, R. N., Montiel, P. O., & Johnstone, K. (2000). Influence of growth temperature on lipid and soluble carbohydrate synthesis by fungi isolated from fellfield soil in the maritime Antarctic. *Mycologia*, 92(2), 222–229. <https://doi.org/10.1080/00275514.2000.12061148>
- Wilson, S. A., & Roberts, S. C. (2012). Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant Biotechnology Journal*, 10(3), 249–268. <https://doi.org/10.1111/j.1467-7652.2011.00664.x>
- Yu, Y., Li, T., Wu, N., Jiang, L., Ji, X., & Huang, H. (2017). The Role of Lipid Droplets in *Mortierella alpina* Aging Revealed by Integrative Subcellular and Whole-Cell Proteome Analysis. *Scientific Reports*, 7(1), 43896. <https://doi.org/10.1038/srep43896>
- Zeng, T., Chen, Y., Jian, Y., Zhang, F., & Wu, R. (2022). Chemotaxonomic investigation of plant terpenoids with an established database (TeroMOL). *New Phytologist*, 235(2), 662–673. <https://doi.org/10.1111/nph.18133>
- Zeng, T., Liu, Z., Liu, H., He, W., Tang, X., Xie, L., & Wu, R. (2019). Exploring Chemical and Biological Space of Terpenoids. *Journal of Chemical Information and Modeling*, 59(9), 3667–3678. <https://doi.org/10.1021/acs.jcim.9b00443>
- Zeng, T., Liu, Z., Zhuang, J., Jiang, Y., He, W., Diao, H., Lv, N., Jian, Y., Liang, D., Qiu, Y., Zhang, R., Zhang, F., Tang, X., & Wu, R. (2020). TeroKit: A Database-Driven Web Server for Terpenome Research. *Journal of Chemical Information and Modeling*, 60(4), 2082–2090. <https://doi.org/10.1021/acs.jcim.0c00141>
- Zerbe, P., & Bohlmann, J. (2014). Bioproducts, Biofuels, and Perfumes: Conifer Terpene Syntheses and their Potential for Metabolic Engineering. In R. Jetter (Ed.), *Phytochemicals – Biosynthesis, Function and Application: Volume 44* (pp. 85–107). Springer International Publishing. [https://doi.org/10.1007/978-3-319-04045-5\\_5](https://doi.org/10.1007/978-3-319-04045-5_5)
- Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R., & Bohlmann, J. (2012). Bifunctional *cis*-Abienol Synthase from *Abies balsamea* Discovered by Transcriptome Sequencing and Its Implications for Diterpenoid Fragrance Production. *The Journal of Biological Chemistry*, 287(15), 12121–12131. <https://doi.org/10.1074/jbc.M111.317669>
- Zhan, X., Zhang, Y.-H., Chen, D.-F., & Simonsen, H. T. (2014). Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and  $\alpha/\beta$ -santalene. *Frontiers in Plant Science*, 5. <https://www.frontiersin.org/articles/10.3389/fpls.2014.00636>

- Zhang, B., Fang, Z., Chen, J., Wu, R., & Mao, B. (2024). Dual flocculation strategy with pH adjustment for cost-effective algae harvesting. *Journal of Water Process Engineering*, 59, 105009. <https://doi.org/10.1016/j.jwpe.2024.105009>
- Zhang, H., Cui, Q., & Song, X. (2021). Research advances on arachidonic acid production by fermentation and genetic modification of *Mortierella alpina*. *World Journal of Microbiology and Biotechnology*, 37(1), 4. <https://doi.org/10.1007/s11274-020-02984-2>
- Zhang, K., Bonito, G., Hsu, C.-M., Hameed, K., Vilgalys, R., & Liao, H.-L. (2020). *Mortierella elongata* Increases Plant Biomass among Non-Leguminous Crop Species. *Agronomy*, 10(5), Article 5. <https://doi.org/10.3390/agronomy10050754>
- Zhao, D.-D., Jiang, L.-L., Li, H.-Y., Yan, P.-F., & Zhang, Y.-L. (2016). Chemical Components and Pharmacological Activities of Terpene Natural Products from the Genus *Paeonia*. *Molecules (Basel, Switzerland)*, 21(10), 1362. <https://doi.org/10.3390/molecules21101362>
- Zhou, F., & Pichersky, E. (2020). More is better: The diversity of terpene metabolism in plants. *Current Opinion in Plant Biology*, 55, 1–10. <https://doi.org/10.1016/j.pbi.2020.01.005>

## CHAPTER 2

### **Multilevel analysis between *Physcomitrium patens* and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi**

This chapter is adapted from its original publication in *The Plant Journal*:

Mathieu, D., Bryson, A. E., Hamberger, B., Singan, V., Keymanesh, K., Wang, M., Barry, K., Mondo, S., Pangilinan, J., Koriabine, M., Grigoriev, I. V., Bonito, G., & Hamberger, B. Multilevel analysis between *Physcomitrium patens* and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi. *The Plant Journal* 2024. <https://doi.org/10.1111/tpj.16605>.

## Abstract

The model moss species *Physcomitrium patens* has long been used for studying divergence of land plants spanning from bryophytes to angiosperms. In addition to its phylogenetic relationships, the limited number of differential tissues, and comparable morphology to the earliest embryophytes provide a system to represent basic plant architecture. Based on plant-fungal interactions today, it is hypothesized these kingdoms have a long-standing relationship, predating plant terrestrialization. Mortierellaceae have origins diverging from other land fungi paralleling bryophyte divergence, are related to arbuscular mycorrhizal fungi but are free-living, observed to interact with plants, and can be found in moss microbiomes globally. Due to their parallel origins, we assess here how two Mortierellaceae species, *Linnemannia elongata* and *Benniella erionia*, interact with *P. patens* in coculture. We also assess how Mollicute-related or Burkholderia-related endobacterial symbionts (MRE or BRE) of these fungi impact plant response. Coculture interactions are investigated through high-throughput phenomics, microscopy, RNA-sequencing, differential expression profiling, gene ontology enrichment, and comparisons among 99 other *P. patens* transcriptomic studies. Here we present new high-throughput approaches for measuring *P. patens* growth, identify novel expression of over 800 genes that are not expressed on traditional agar media, identify subtle interactions between *P. patens* and Mortierellaceae, and observe changes to plant-fungal interactions dependent on whether MRE or BRE are present. Our study provides insights into how plants and fungal partners may have interacted throughout history based on their communications observed today and identify *L. elongata* and *B. erionia* as modern fungal endophytes with *P. patens*.

## **Keywords**

*Physcomitrium patens*, Mortierellaceae, endobacteria, RNA-sequencing, differential expression, gene ontology enrichment, RaspberryPi, PlantCV

## **Significance Statement**

We implement high resolution automated phenomics, bright field microscopy, and high throughput transcriptomic analysis with the moss, *Physcomitrium patens*, in coculture with fungal endophytes in Mortierellaceae, *Linnemannia elongata* and *Benniella erionia* either containing or lacking intracellular bacterial symbionts. The nature of this interaction represents an untested facet of plant-fungal-bacterial relations and may give insight into how long standing evolutionarily conserved traits influencing cross-Kingdom communication have been retained.

## **Introduction**

Plants and fungi have a long history of symbiosis and cohabitation, with over 90% of modern land plants demonstrating some degree of mutualism [Bonfante & Genre 2010, Smith & Read 2010]. In addition to the high frequency of plant-fungal interaction among land plants, the observed mutualism extending to algae and lichens implies that the emergence of traits allowing for a beneficial exchange of compounds between plant and fungus arose even earlier in chlorophyllic phototroph evolution [Du et al. 2019, Duckett et al. 2006, Knack et al. 2015, Kohler et al. 2015, Hanke & Rensing 2010, Liepina 2012, Loron et al. 2019, Lutzoni et al. 2001, Lutzoni et al. 2018, Morris et al. 2018, Nelson et al. 2019, Russell & Bulman 2005]. Although bacteria and fungi are known to have dominated the terrestrial landscape long before plant terrestrialization, the ability for plant and fungal kingdoms to interact early in embryophyte evolution may have enabled the global takeover of both Kingdoms. Today, this relationship is exemplified through plants exchanging a reliable carbon source via the products of photosynthesis (i.e., sugars and fatty acids)

and nearly ubiquitous symbiosis with filamentous fungi that exchange nitrogen, phosphorous, micronutrients, metabolites and water retention [Bonfante & Genre 2010, Martin & Nehls 2009]. Further, modern plant and fungal symbionts have been shown to mitigate many shared stresses such as oxidative, osmotic, heat, UV radiation, and rapid temperature flux [de Vries & Archibald 2018, Du et al. 2019, Fürst-Jansen et al. 2020, Jermy 2011, Kohler et al. 2015, Lutzoni et al. 2018]. These same stresses would have posed significant barriers to entry for the first terrestrial land plants as well. The early emergence of plant-fungal interactions may have reduced constraints imposed by the ancient terrestrial landscape, and consequently may have led to the global expansion of plants and fungi observed today.

While many plant-fungal mutualists have been identified in embryophytes, no reports of fungal mutualism in the model moss *Physcomitrium patens* (formerly *Physcomitrella patens*) have been made [Bonfante & Genre 2010, Read et al. 2000]. This is despite many arbuscular mycorrhizal fungi that have demonstrated capacity for mutualism in other bryophytes like hornworts and liverworts [Fonseca & Berbara 2008, Ligrone et al. 2007]. *P. patens* is capable of specialized fungal response although this is largely in the context to combatting parasitic fungi which otherwise would decrease host fitness [Bressendorff et al. 2016, Davey et al. 2009, Delaux & Schornack 2021, Lehtonen et al. 2009, Lehtonen et al. 2012, Mittag et al. 2015, Ponce de Leon 2011, Ponce de Leon et al. 2012]. Additionally, evidence that *P. patens* has (or had) the capacity for interacting with fungi can be supported by the presence of orthologs essential to detecting and forming plant-fungal interaction. Some conserved genes that are indicative of this possibility include a chitin like receptor *PpCERKI* necessary to signal environmental presence of fungi, a *VAPYRIN* like homolog with only known function in forming symbiotic interaction between plants and fungi, and functional strigolactone hormone pathways with secondary functions

known to signal host root proximity to symbiotic and parasitic fungi [Bressendorff et al. 2016, Delaux & Schornack 2021, Proust et al. 2011, Rathgeb et al. 2020].

Here we investigate two filamentous fungal species belonging to Mortierellaceae as potential symbiotic candidates with *P. patens*. Mortierellaceae are a lineage of free-living fungi closely related to arbuscular mycorrhizal fungi, that are known to improve aboveground plant growth and development and to associate with plants as endophytes [Vandepol et al. 2022, Johnson et al. 2019, Zhang et al. 2021]. Fungi in Mortierellaceae embody many promising traits as a mutualist, sharing an evolutionary history with the widespread but host-dependent arbuscular mycorrhizal fungi (AMF) and forming mutualisms with chlorophytes (algae), *Arabidopsis thaliana*, and other embryophytes [Becker & Cuta 2020, Du et al. 2019, Johnson et al. 2019, Rensing et al. 2008, VandePol et al. 2022, Zhang et al. 2021]. Interestingly, species in both AMF and Mortierellaceae can be colonized by either Mollicute-related endobacteria (MRE) or Burkholderia-related endobacteria (BRE), which grow within host cells and are nutritionally dependent on the host. We investigated interactions of *P. patens* with *Linnemannia elongata* (formerly *Mortierella elongata* strain NVP64) and *Benniella erionia* (formerly strain GB\_Aus27b), either carrying (WT) or cleared (CU) of its bacterial endosymbiont (described in more detail below). There are multiple reports of *L. elongata* forming mutualistic interactions with algae and plants in ways that increase plastid size, aboveground plant growth, flowering, and seed production in different plant species [Du et al. 2019, VandePol et al. 2022]. In contrast, the recently described fast-growing fungus *B. erionia* caused chlorosis in interactions with the algal species *Nannochloropsis oceanica* and *Chlamydomonas reinhardtii* [Du et al. 2019]. The lack of identified interaction of *P. patens* with AMF, which is the most widespread embryophytic mutualist seen today [Feijen et al. 2018], may be due to AMF predominantly colonizing roots, a tissue absent in moss. Despite the current global

abundance of AMF, ancestral reconstruction suggests that the prolificity seen today parallels the emergence and expansion of angiosperms 250 MYA, while the fungal species detected during plant terrestrialization more closely resembles Mortierellaceae [Fiejen et al. 2018].

We also assessed the impact of endobacterial symbionts of fungi on fungal-moss interactions. Previous studies have shown that MRE and BRE intracellular bacteria can be removed from the hosts with antibiotics, which results in changed fungal growth and metabolism [Desirò et al. 2017, Uehling et al. 2017, VandePol 2022]. As *L. elongata* naturally contains BRE and *B. erionia* naturally contains MRE, we carried out our experiment using isogenic isolates either with (WT) or without (CU) endosymbionts. Previous studies with AMF have found that BRE increases sporulation in their host and improves energy capacity/availability, often at the expense of a reduced growth rate [Alabid et al. 2019, Salvioli et al. 2016, Uehling et al. 2017]. While the exact impact that endobacteria have specifically on *L. elongata*, *B. erionia*, and their plant associates are unclear, endobacteria are known to influence how fungi interact with their environment, and therefore may play an important role in plant-fungal interactions [Desirò et al. 2017, Guo & Narisawa 2018, Ohshima et al. 2016, Uehling et al. 2017, VandePol 2020].

To measure the interaction between fungal endophytes and *P. patens*, we investigated the interaction at organismal, cellular, and transcriptional levels. This was accomplished through a custom built phenomics platform and analysis pipeline, brightfield microscopy, and RNA-sequencing with subsequent expression analysis. Broader influences were also investigated through incorporation of transcriptional analysis data and comparison with the results from the ‘*Physcomitrella patens* Gene Atlas Project’ (‘Gene Atlas Project’), which examined 99 RNA-seq expression datasets generated with *P. patens* transcripts [Perroud et al. 2018]. Our results indicate distinct responses in *P. patens* when cocultured with *B. erionia* or *L. elongata* and the nature of

their response being dictated by endobacteria. Here, we propose that *P. patens* has retained some ability to interact with Mortierellaceae endophytically based on the observation of asymptomatic intracellular colonization of fungi within plant tissues and these colonizations potentially being dependent upon the presence of endobacteria. These observations provide insights for an interaction that may have originated 500 MYA [Fiejen et al. 2018, Hobbie & Boyce 2010, Ivarsson et al. 2020].

## Methods & Materials

### ***Bright field microscopy of P. patens and Mortierellaceae in coculture***

*P. patens* (Gransden 2004; strain Pp40001) and *L. elongata* wildtype (WT), *L. elongata* cured of BRE (CU), *B. erionia* wildtype (WT) and *B. erionia* cured of MRE (CU) were cocultured on opposite halves of BCD agar media (1mM MgSO<sub>4</sub>, 1.84 mM KH<sub>2</sub>PO<sub>4</sub>, 10 mM KNO<sub>3</sub>, 12.5mg FeSO<sub>4</sub>·7H<sub>2</sub>O, 7g Agar, 1mL Hoagland's A-Z trace element solution, H<sub>2</sub>O to 1L) for two weeks [Ashton & Cove 1977, Wang & He 2015]. *B. erionia* and *L. elongata* were cured of endobacteria in previous work by cycling between liquid and agar medias for one-week intervals over fifteen weeks with a combination of four antibiotics (80 µg/ml Ampicillin, 50 µg/ml Kanamycin, 50 µg/ml Streptomycin, and 120 µg/ml Ciprofloxacin) [Desirò et al. 2017, Ueling et al. 2017]. The resulting fungal strains were then confirmed for the absence of endobacteria through sequencing. Whole plant-fungal cocultures, which mainly consisted of *P. patens* protonema, and to a lesser extent *P. patens* rhizoid and leaflets (all asexual tissues) along with fungal hyphae, were collected by scraping all tissue off the surface of the plate with tweezers, placing in 1.5 mL Eppendorf tubes, and cleared with 1 mL formalin-aceto-alcohol (FAA) solution (50 Ethanol: 5 Glacial Acetic Acid: 10 Formalin: 35 H<sub>2</sub>O). The cocultures were then placed under a vacuum for 30 minutes and stored

in the dark overnight. The next day, FAA solution was removed with a pipette, then tissue was stained with 1 mL 1% Chlorazol Black E for 24 hours. Samples were destained with 25%, 50%, 75%, Ethanol solutions each for 20 minutes, then stored in 100% Ethanol solution until microscopy. Samples were then placed on microscope slide with glycerol and viewed with a bright field microscope (Model: Leica DM750) (Figure 2.2; Supplemental Data 1).

### ***Automated phenotyping of P. patens and Mortierellaceae using RaspberryPi***

*P. patens* (Gransden 2004; strain Pp40001) was inoculated on 100g wet Redi-earth soil in 1-liter, wide-mouth mason jars, under fluorescent lights, on top of a black tarp, with white walls surrounding the experiment. A sterilized, white, metal thumbtack was placed in the center of the soil for each sample to correct for white balance. Samples were watered every 14 days (or as needed to keep soil damp), with 10mL distilled water. One-piece, twist-top, wide mouth lids were used and had two, 1cm diameter holes drilled, for camera placement and watering, respectively. *P. patens* in BCD-agar were blended until homogenous at the day of inoculation, and evenly dispersed among samples. The four fungal strains were grown independently for inoculation on 2g of perlite with 10mL of minimal malt extract (5g/L malt extract, 0.25g/L yeast extract) liquid media for two weeks. Fungal treatments were randomly selected for inoculation to jars following the two-week growth phase. In each jar 2g of saturated, colonized perlite was dispersed to their respective treatment (10 control, 8 *P. patens* x *B. erionia* WT, 8 *P. patens* x *L. elongata* WT, 7 *P. patens* x *B. erionia* CU, 7 *P. patens* x *L. elongata* CU). The control treatment was inoculated with 2g of perlite and sterile malt extract media.

Monitoring of growth occurred with RaspberryPi Microcomputers 3 ModelB V1.2, which were initialized following online instructions<sup>2</sup> and all devices were fitted with an Arducam Multi

---

<sup>2</sup> <https://projects.raspberrypi.org/en/projects/raspberry-pi-setting-up>

Camera Adapter Module V2.1 with four Raspberry Pi V2.1 Cameras. Scripts to run multi-camera adapter and camera were downloaded and set up following instructions for this specific RaspberryPi hardware<sup>3</sup>. Images were automatically captured daily at 12:00 PM EST (Supplemental Data 2a). Double sided 2x2cm scotch foam tape with 8mm diameter cut hole was placed on jar lids to secure cameras with a perpendicular view ~15cm away from each sample then secured to position with tape.

### ***Image analysis using PlantCV***

Images were annotated to include key information, i.e., their origin and time of image capture Cam#\_lens\_#\_YEAR-MM-DD\_HH-MM.jpg (ex: Cam1\_lens\_1\_2021-08-13\_12-00.jpg). Python script (Supplemental Code 1) for quantifying green pixels in each image ran in Jupyter Notebook [Kluyver et al. 2016]. PlantCV installation<sup>4</sup> was done following Anaconda<sup>5</sup> specific instructions [Berry et al. 2018, Gehan et al. 2017]. Necessary dependencies were imported for analysis (os [v3.7.6], numpy [v1.18.1], cv2 [v3.4.9], matplotlib [v3.1.3], plantcv [v3.8.0], pandas [v1.0.1], glob [v3.7.6]). Each picture was then loaded and labeled based on treatment, date, computer, and camera. The white thumbtack centered in each mason jar was used to correct for white balance. Each image was converted from RGB (red, green, blue) to the LAB (lightness, magenta/green, blue/yellow) color space. The threshold for plant health was designated as anything in the pixel value range among the “A” (green-magenta) values between 121-255. This range was selected due to it distinguishing only healthy, green moss tissue. Clusters of pixel groups less than 100 were removed and dilated to reduce noise. Final appended counts along with treatment, date, computer,

---

<sup>3</sup> [https://www.arducam.com/downloads/RaspCAM/RaspberryPi\\_Multi\\_Camera\\_Adapter\\_Module\\_UG.pdf](https://www.arducam.com/downloads/RaspCAM/RaspberryPi_Multi_Camera_Adapter_Module_UG.pdf)

<sup>4</sup> <https://plantcv.readthedocs.io/en/latest/installation/>

<sup>5</sup> <https://www.anaconda.com/>

and camera were then exported as a .csv for data visualization and evaluation (Supplemental Data 2b).

### ***Propagation of P. patens in coculture with Mortierellaceae species for RNA extraction***

Samples were grown on autoclaved Redi-earth mix, inoculated with liquid BCD media in sterile glass jars with 0.5-micron pore vents under 50  $\mu\text{mol}$  LED lights over the course of 35 days with fungal inoculation occurring on day 10. The 25 days spent in coculture provided time for fungi and plants to sufficiently overcome transplant stresses and naturally colonize the soil together (to avoid differential expression and variation that may be caused solely by changing environments). The five experimental conditions were grown in triplicate with *P. patens* (Genotype: Gransden 2004; strain: Pp40001) grown in seclusion, cocultured with *L. elongata* strains containing or cured of endobacteria (WT/CU), and cocultured with *B. erionia* strains containing or cured of endobacteria (WT/CU). Fungi in isolation were also cultivated, however, tissue retrieved was insufficient for RNA sequencing. Using tweezers, samples were extracted and separated from the soil 25 days after inoculation, removing growth substrate while maintaining *P. patens* rhizoid and fungal hyphae structures. Samples were flash frozen in liquid N<sub>2</sub> in 2mL Eppendorf tubes and stored at -80° C until RNA extraction.

### ***RNA extraction and quality control for Illumina library preparation***

A phenol-chloroform based RNA extraction was performed as described in previous work [Kolossova et al. 2004] but adapted to 100mg of plant tissue. After extraction, 1  $\mu\text{L}$  RNA was run on a 1% agarose gel to confirm integrity. Samples were treated with DNase following manufacturer's instructions (ThermoFisher, Product AM1907). Submitted RNA was quality checked with an Agilent 2100 Flowcell Bioanalyzer (RNA integrity value  $8.11 \pm 0.15$  (95% CI)), DeNovix DS-11 Nanodrop ( $273.8 \pm 41.3$  ng/ $\mu\text{L}$  (95% CI)), and Qubit 2.0 Fluorometer.

### ***RNA sequencing, processing, alignment, and data management***

Plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the protocol outlined by Illumina in their user guide<sup>6</sup> and with the following conditions: total RNA starting material was 1 µg per sample and 8 cycles of PCR was used for library amplification. The prepared libraries were then quantified using KAPA Illumina library quantification kit (Roche) and run on a LightCycler 480 real-time PCR instrument (Roche). The quantified libraries were then multiplexed, and the pool of libraries was then prepared for sequencing on the Illumina NovaSeq 6000 sequencing platform using NovaSeq XP v1 reagent kits (Illumina), S4 flow cell, following a 2x150 indexed run recipe [Modi et al. 2021].

We generated a total of 664 million paired reads (per sample: 49M ± 3.9M (95% CI)) and data is available on the JGI online genome database (gold.jgi.doe.gov/projects; GOLD Project ID Gp0332982-Gp0332996) and the NCBI SRA database (PRJNA807682). All reads were mapped back to *P. patens* genome version 3.3 CDS file 2018 [Ppatens\_318\_V3.3.cds.fa.gz; Phytozome; NCBI Taxonomy ID: 3218] (*P. patens* transcriptome) [Lang et al. 2018], Filtered *Mortierella* GBAus27b CDS file [MorGBAus27b\_1\_GeneCatalog\_CDS\_20170422.fa.gz, MycoCosm; NCBI Taxonomy ID: 1954212] (*B. erionia* transcriptome) [Chang et al. 2022], and filtered *Mortierella* NVP64 CDS file [MoeNVP64\_1\_GeneCatalog\_CDS\_20190403.fa.gz; MycoCosm; NCBI Taxonomy ID: 2684331] (*L. elongata* transcriptome). Prior to mapping, meta-transcriptomes were created representing their respective environmental conditions by concatenating all genes from *P. patens* with either the *B. erionia* transcriptome or *L. elongata* transcriptome when applicable. Supplemental bash script (Supplemental Code 2a-b) was submitted to the Michigan State

---

<sup>6</sup> [https://support.illumina.com/sequencing/sequencing\\_kits/truseq-stranded-mrna.html](https://support.illumina.com/sequencing/sequencing_kits/truseq-stranded-mrna.html)

University High Performance Computing Center (MSU HPCC). The transcriptome and meta-transcriptomes (*P. patens*; *P. patens* x *B. erionia*; *P. patens* x *L. elongata*) were indexed using Salmon [v1.2.1] and the ‘salmon index’ function to create reference libraries for downstream read quantification. Raw reads were split into forward and reverse read files, raw quality control (QC) reports were generated with FastQC [v0.11.7], low-quality reads and adapters were removed using FastP [v0.21.0] which maintained  $97.33 \pm 0.64\%$  (95% CI) of the original sequence [Andrews 2010, Chen et al. 2018, Patro et al. 2017]. Trimmed QC reports were again generated with FastQC, and filtered reads were mapped to their respective index using the ‘salmon quant’ function [Andrews 2010]. In total there were five treatments, each in triplicate (JGI Sample Barcode; *P. patens*: CUTON, CUTOO, CUTOP; *P. patens* x *B. erionia* WT: CUTOS, CUTOT, CUTOU; *P. patens* x *B. erionia* CU: CUTOW, CUTOX, CWYAU; *P. patens* x *L. elongata* WT: CUTOZ, CUTPA, CUTPB; *P. patens* x *L. elongata* CU: CUTPC, CUTPG, CUTPH)) (Supplemental Data 3a).

### ***Differential gene expression analysis using DESeq2***

Expression of *P. patens* genes from each treatment were differentiated and determined as significant using DESeq2 [v1.26.0] (Supplemental Code 3) [Love et al. 2014]. The *P. patens* mapped read dataset (Supplemental Data 3a) was used as an input after formatting with tximport [v1.22.0], to correct for multiple isoforms across samples [Soneson et al. 2016]. The algorithm modeled relative library depth, dispersion of individual gene counts, and significance of coefficients, all of which were used to determine library size and dispersion corrected negative binomial general linearized model. Genes were considered significant if they demonstrated high differential expression differences based on the Bonferroni adjusted P-value ( $P_{adj} < 0.01$ ) (Supplemental Data 3b-e).

### ***Comparative analysis across the ‘Gene Atlas Project’ and newly tested conditions here***

The quantified reads per kilobase per million (RPKM) datasets from ‘The *Physcomitrella patens* Gene Atlas Project’ were downloaded to compare our 15 samples to their 99 *P. patens* samples representing various tissues, treatments, growth stages, and laboratory biases to untangle globally shared trends in expression [Perroud et al. 2018]. RPKM values were converted to transcript per million (TPM) using the following equation on the dataset:  $TPM = \frac{RPKM}{\sum RPKM} * 10^6$  (Supplemental Data Xa-b; Supplemental Code 4) because TPM is a unitless metric and resembles a percentage based system, it corrects for varying library sizes, and serves as a better tool for inter-sample comparison [Zhao et al. 2020]. The dataset was simplified by removing genes with summed expression less than 20 TPM across all samples ( $\sum_{114} TPM < 20$ ), which reduced the gene dataset by 51.47% (16256/31581). Differential expression of samples between results reported here and within the ‘Gene Atlas Project’ were compared using DESeq2 (Supplemental Data Xc; Supplemental Code 4) [Love et al. 2014]. A principal component analysis was performed to identify clustering based on expression to identify similarities or differences between our samples and the ‘Gene Atlas Project’. Additionally, comparisons were made across all samples with  $\sum_{99} TPM = 0$  among all ‘Gene Atlas Project’ samples and then compared to our samples where  $\sum_{15} TPM \neq 0$  to identify any novel gene expression (Supplemental Data Xd). This comparison was also made in reverse to identify which genes were silenced in our samples, in which  $\sum_{15} TPM = 0$  and  $\sum_{99} TPM \neq 0$  among the ‘Gene Atlas Project’ (Supplemental Data Xe).

### ***Analysis of transcriptional differences in *P. patens* based on growth media effects***

Comparisons based on common *P. patens* media sampled 41 libraries from gametophore tissues grown on BCD Agar (5), BCDAT (6), Hoagland (6), Knop agar (8), Knop liquid (8), PpNH4 protoplast solution (5) and soil (3). This was taken from 5 different projects (NCBI:

PRJNA751102, PRJNA880579, PRJNA723997, PRJNA259147, PRJNA807682) (Supplemental Data 5) [Causier et al. 2023, Garcias-Morales et al. 2021, Otero et al. 2021, Perroud et al. 2018]. Select SRA “control” reads from each project were downloaded using SRA-toolkit [v3.0.3], trimmed with Fastp [v0.21.0], the *P. patens* transcriptome was again indexed and used for mapping reads with Salmon [v1.8.0], and finally quantified read files were merged using the salmon merge function (Supplemental Code 5). PCA was generated as before using Supplemental Code 4, but with the media-based quant file as an input and classification of each variable based on media (Supplemental Data 5; Supplemental Code 5).

### ***M. elongata* NVP64 genome and transcriptome sequencing, assembly & annotation**

The *Mortierella elongata* NVP64 genome was sequenced using the PacBio Sequel platform. 5 µg of genomic DNA was sheared to >10kb using Covaris g-Tubes. The sheared DNA was treated with exonuclease to remove single-stranded ends and DNA damage repair mix followed by end repair and ligation of blunt adapters using SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). The library was purified with AMPure PB beads. PacBio Sequencing primer was then annealed to the SMRTbell template library and sequencing polymerase was bound to them using Sequel Binding kit 3.0. The prepared SMRTbell template libraries were then sequenced on a Pacific Biosciences's Sequel sequencer using v3 sequencing primer, 1M v3 SMRT cells, and Version 3.0 sequencing chemistry with a 1x360 sequencing movie run time. Filtered subread data was assembled with Falcon version (pb-assembly 0.0.2, falcon-kit 1.2.3, pypeflow 2.1.0) (<https://github.com/PacificBiosciences/FALCON>) to generate an initial assembly and genome statistics (Supplemental Data 6; NCBI Accession: JAXBDG000000000). Mitochondrial sequence was assembled separately from the Falcon pre-assembled reads (reads) using an in-house tool (assemblemito.sh), used to filter the reads, and polished with Arrow version

SMRTLink [v6.0.0.47841] (<https://github.com/PacificBiosciences/GenomicConsensus>). A secondary Falcon assembly was generated using the mitochondria-filtered reads, improved with finisherSC [v2.1] [Lam et al.,2014], and polished with Arrow version SMRTLink [v6.0.0.47841] (<https://github.com/PacificBiosciences/GenomicConsensus>). Completeness of the euchromatic portion of the genome assembly was assessed by aligning assembled consensus RNA sequence data with bbtools [v38.41] bbmap.sh [k=13 maxindel=100000 customtag ordered nodisk] and bbest.sh [fraction=85] (<http://sourceforge.net/projects/bbmap>). Contigs less than 1000 bp were excluded.

We additionally sequenced and assembled a de-novo transcriptome for *M. elongata* NVP64, which provided RNA evidence for improved gene calling. Plate-based RNA sample prep was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Illumina's TruSeq Stranded mRNA HT sample prep kit utilizing poly-A selection of mRNA following the protocol outlined by Illumina in their user guide ([http://support.illumina.com/sequencing/sequencing\\_kits/truseq\\_stranded\\_mrna\\_ht\\_sample\\_prep\\_kit.html](http://support.illumina.com/sequencing/sequencing_kits/truseq_stranded_mrna_ht_sample_prep_kit.html)) and with the following conditions: total RNA starting material was 1 µg per sample and 8 cycles of PCR was used for library amplification. The prepared library was quantified using KAPA Biosystem's next-generation sequencing library qPCR kit (Roche) and run on a Roche LightCycler 480 real-time PCR instrument. The quantified library was then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq 2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2x150 indexed run recipe. Raw fastq file reads were filtered and trimmed using the

JGI QC pipeline resulting in the filtered fastq file. Using BBDuk<sup>7</sup>, raw reads were evaluated for artifact sequence by kmer matching (kmer=25), allowing 1 mismatch and detected artifact was trimmed from the 3' end of the reads. RNA spike-in reads, PhiX reads and reads containing any Ns were removed. Quality trimming was performed using the phred trimming method set at Q6. Finally, following trimming, reads under the length threshold were removed (minimum length 25 bases or 1/3 of the original read length - whichever is longer). Filtered fastq files were used as input for de novo assembly of RNA contigs. Reads were assembled into consensus sequences using Trinity [v2.3.2] [Grabherr et al. 2011]. Trinity was run with the --normalize\_reads (In-silico normalization routine) and --jaccard\_clip (Minimizing fusion transcripts derived from gene dense genomes) options. Incorporating this de-novo transcriptome and filtered RNAseq reads, the *M. elongata* NVP 64 genome was then annotated using the JGI annotation pipeline [Grigoriev et al., 2014].

### ***Orthologous differential expression analysis of P. patens***

Genes identified as essential for plant-fungal symbiosis and their homologs were found in previous work including those of *P. patens* [Delaux et al. 2015]. Homologous gene models from this work were extracted and updated to the newer *P. patens* gene models [V3.3] from the latest assembly and were separated based on gene function. These genes were then searched within DGE analysis of each *P. patens*-Mortierellaceae coculture (Supplemental Data 3b-e) and, if hit, were interpreted as up- or down- regulated among that condition (Supplemental Data 7).

Reads of *L. elongata* (WT/CU) coculture with *A. thaliana* were retrieved from the SRA database (PRJNA704083; Supplemental Data 8a) and with *C. reinhardtii* are now available on the SRA database (PRJNA809543; Supplemental Data 8a). *A. thaliana* CDS (Araport11\_cds\_20220103.gz)

---

<sup>7</sup> <https://sourceforge.net/projects/bbmap/>

and *C. reinhardtii* CDS (Chlamydomonas\_reinhardtii.Chlamydomonas\_reinhardtii\_v5.5-.cds.all.fa.gz) were obtained from The Arabidopsis Information Resource (TAIR; arabidopsis.org) and EnsemblePlants (plants.ensembl.org) respectively. Reads were quantified and analyzed in the same pipeline previously described, using FastQC [v0.11.7], FastP [v0.21.0], Salmon [v1.2.1], tximport [v1.22.0], DESeq2 [v1.26.0], and g:Profiler (Supplemental Data W, Vab; Supplemental Code 6a-c). Orthofinder [v2.5.2] was used to identify single copy orthologs between plant systems [Emms & Kelly 2019]. Amino acid sequences were obtained through Ensembl (Arabidopsis\_thaliana.TAIR10.pep.all.fa.gz, Chlamydomonas\_reinhardtii.Chlamydomonas\_reinhardtii\_v5.5.pep.all.fa.gz, Physcomitrium\_patens.Phypha\_V3.pep.all.fa.gz; plants.ensembl.org). Single copy orthogroups (Supplemental Data 8c-f) were manually filtered for overlapping DEGs between *P. patens*, *A. thaliana*, two *C. reinhardtii* treatments (Supplemental Data 8g). This was repeated exclusively between *P. patens* vs. *A. thaliana* and *P. patens* vs. *C. reinhardtii* DEGs (Supplemental Data 8c). The results from this analysis did not demonstrate any clear shared response.

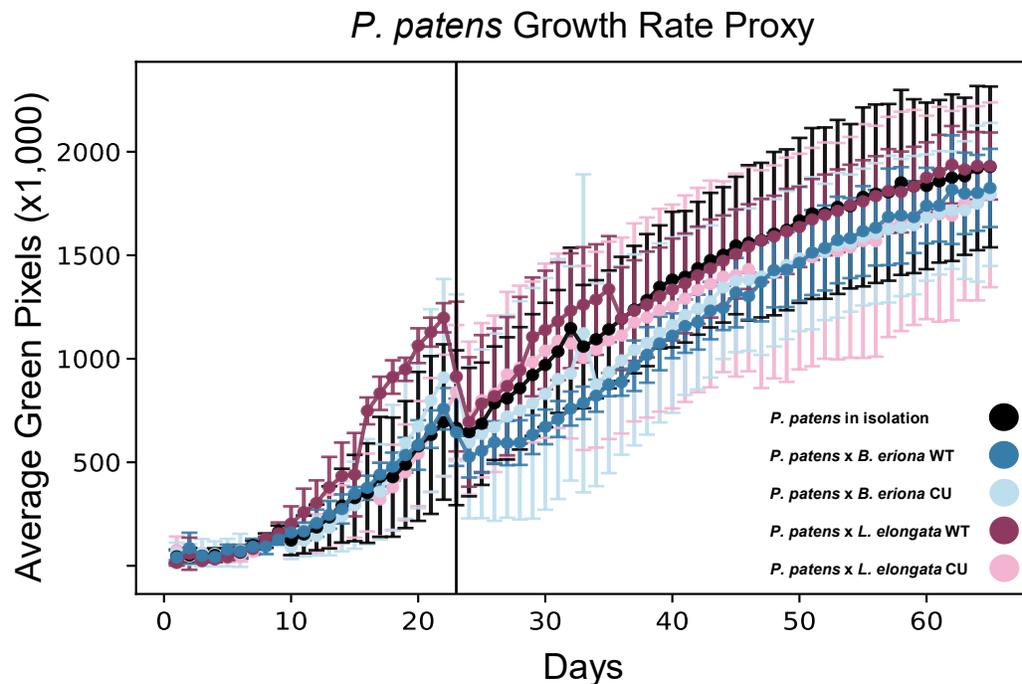
### ***Gene ontology enrichment analysis***

Gene ontology (GO) enrichment was categorized using g:Profiler by submitting the separated list of significant up- or down-regulated genes generated with DESeq2 (Supplemental Data 3b-e) for each experimental condition (Supplemental Data 9). The g:Profiler software was used to identify significantly differentiated GO terms for the four fungal treatments and *P. patens* and overrepresented terms between our samples and the ‘Gene Atlas Project’ [Raudvere et al. 2019, Reimand et al. 2007].

## Results & Discussion

### ***Development and implementation of high-throughput phenomics approach monitors *P. patens* growth***

We have created and utilized a novel phenomics platform which can automatically capture *P. patens* growth over time with exceptional sensitivity and throughput. Over the course of two months and 40 samples, we generated over 2,000 time-stamped images. These images were quantified with resolution indicating growth changes at the millimeter scale (Supplemental Data 2a-b). This was accomplished by using RaspberryPis for image capture and the software PlantCV to quantify plant growth. While sample germination differed between replicates, no systematic advantage or penalty to health was observed for either *P. patens* or fungal partners. Hence, the results at the scale and macroscopic resolution of this experiment do not generate decisive results regarding either negative or positive effects on growth (Figure 2.1). This phenomics approach also provided additional benefit in confirming the cohabitation of both species throughout the entirety of the growth period as seen by the maturing fungi and *P. patens* cultures present throughout the photos (Supplemental Data 2a).



**Figure 2.1:** Daily quantified growth of *P. patens* over 65 days compared to samples inoculated with either *B. erionia* WT, *B. erionia* CU, *L. elongata* WT, or *L. elongata* CU

Average daily green pixel count with standard deviation of *P. patens* grown in isolation (black) or grown in coculture with *B. erionia* WT (dark blue), *B. erionia* CU (light blue), *L. elongata* WT (magenta), *L. elongata* CU (pink). All samples grew without fungi until day of inoculation (Day 23) indicated by vertical black line, in which fungal inoculated (or uninoculated) perlite was added to their respective samples predetermined by a random number generator. The dip in growth on day of inoculation can be attributed to perlite covering up already grown moss tissue.

### ***Microscopy reveals P. patens responds differently and uniquely to each fungal strain***

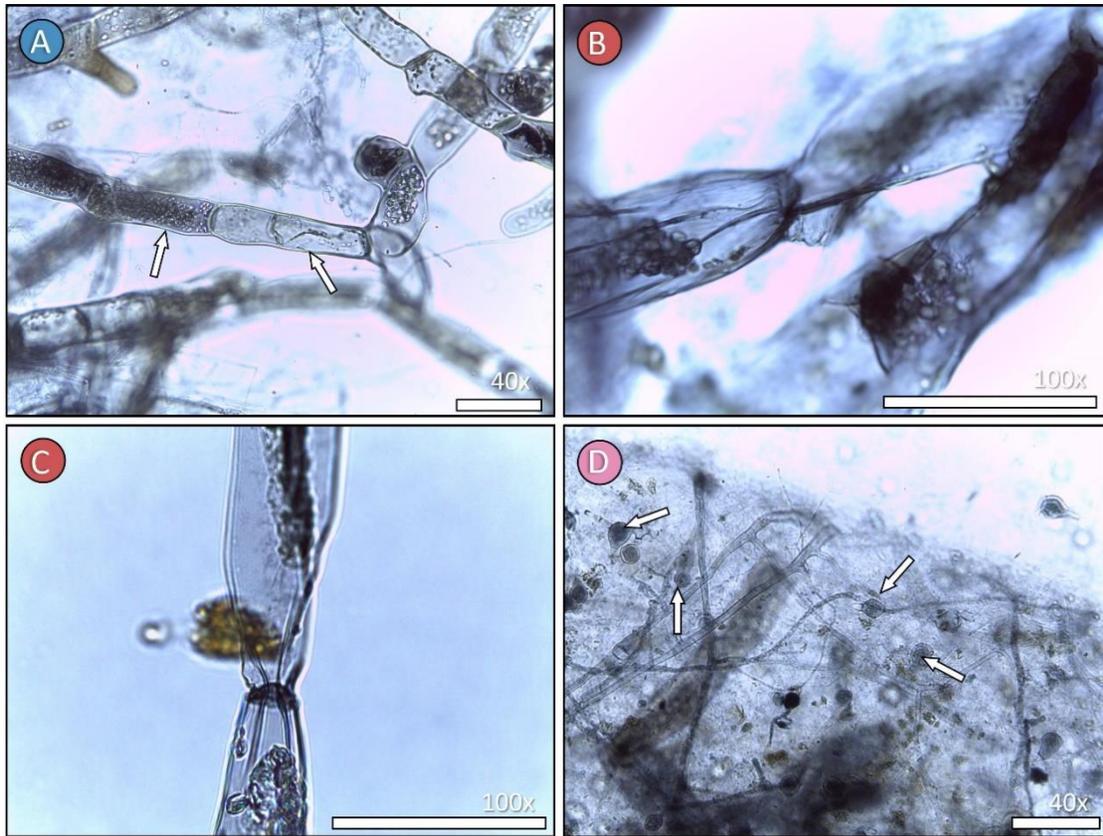
To investigate the physical interaction of fungal species with *P. patens* at a cellular resolution, combinations were subjected to histology. Fungal hyphae were often identified near or within ruptured plant cells, suggesting all fungal samples may have saprotrophic tendencies.

Complementing our phenomic observations, all cocultures showed successful cohabitation with clear maturation and growth of both plant and fungi on BCD agar. *B. erionia* WT cocultures had the most common occurrences of fungal hyphae inhabiting ruptured *P. patens* cells and the highest abundance of hyphae within plant cells. Additionally, we found multiple characteristic

instances of *B. erionia* WT hyphae inside *P. patens* where the plant cell had also retained turgor pressure (Figure 2.2A; Supplemental Data 1). Representative colonized cells experienced bleaching and lacked any observable chloroplasts, with neighboring *P. patens* cells containing an abnormally high density of chloroplasts (Figure 2.2A; Supplemental Data 1). Infected *P. patens* cells were typically limited to a single cell without passage through cellular junctions, which could indicate successfully suppressed infection by *P. patens*. In contrast, *P. patens* and *B. erionia* CU cocultures showed no visible interaction, indicating the potential relevance of endobacteria for colonization. While the exact function(s) of endobacteria in fungi is still speculative, they appear to confer a higher resiliency to environmental conditions and better fungal germination rate in some cases, with evidence here suggesting in some cases endobacteria may also have an influential role in fungal colonization in plant hosts [Naumann et al. 2010, Salvioli et al. 2016].

*Linnemannia elongata* WT cocultures also exhibited intracellular colonization, however *P. patens* retained cellular chloroplast content within fungal inhabited cells (Figure 2.2B & 2.2C; Supplemental Data 1). Unlike *B. erionia* WT, *L. elongata* WT hyphae were observed to cross *P. patens* cellular junctions (Figure 2.2C; Supplemental Data 1). The retention of chloroplasts, the spread of intracellular hyphae, and the retention of turgor pressure in *P. patens* with *L. elongata* WT indicates a less intrusive interaction. *L. elongata* CU uniquely inhabits its environment compared to all other samples by producing a high abundance of chlamydospores (Figure 2.2D; Supplemental Data 1) [Nguyen et al. 2019]. Because of this unique environmental colonization, it is possible that *L. elongata* CU induces a different and unique *P. patens* response compared to the other fungal cocultures. Based on the lack of colonization of *P. patens* from both cured

Mortierellaceae strains and colonization from both WT, it appears that endobacteria may be a critical component for both *L. elongata* and *B. erionia* to interact endophytically.



**Figure 2.2:** Representative and specific interaction of *P. patens* with *B. erionia* and *L. elongata* (White bar indicates 50  $\mu\text{m}$  for each picture)

**A)** Two unruptured *P. patens* protonema cells, one of which (right arrow) has experienced major chlorosis and appears to have a *B. erionia* hyphae encapsulated within the cell, while the neighboring *P. patens* cell (left arrow) has an abnormally high abundance of chloroplasts. **B)** *L. elongata* WT hyphae spanning intercellularly between two *P. patens* cells **C)** Second occurrence of *L. elongata* WT colonizing *P. patens* cells with *L. elongata* WT bridging the gap between *P. patens* cellular junction points **D)** Abundance of *L. elongata* CU sporing bodies in cell culture.

***Comparative transcriptomics indicates distinct response in P. patens by L. elongata, B. erionia, and endobacterial presence***

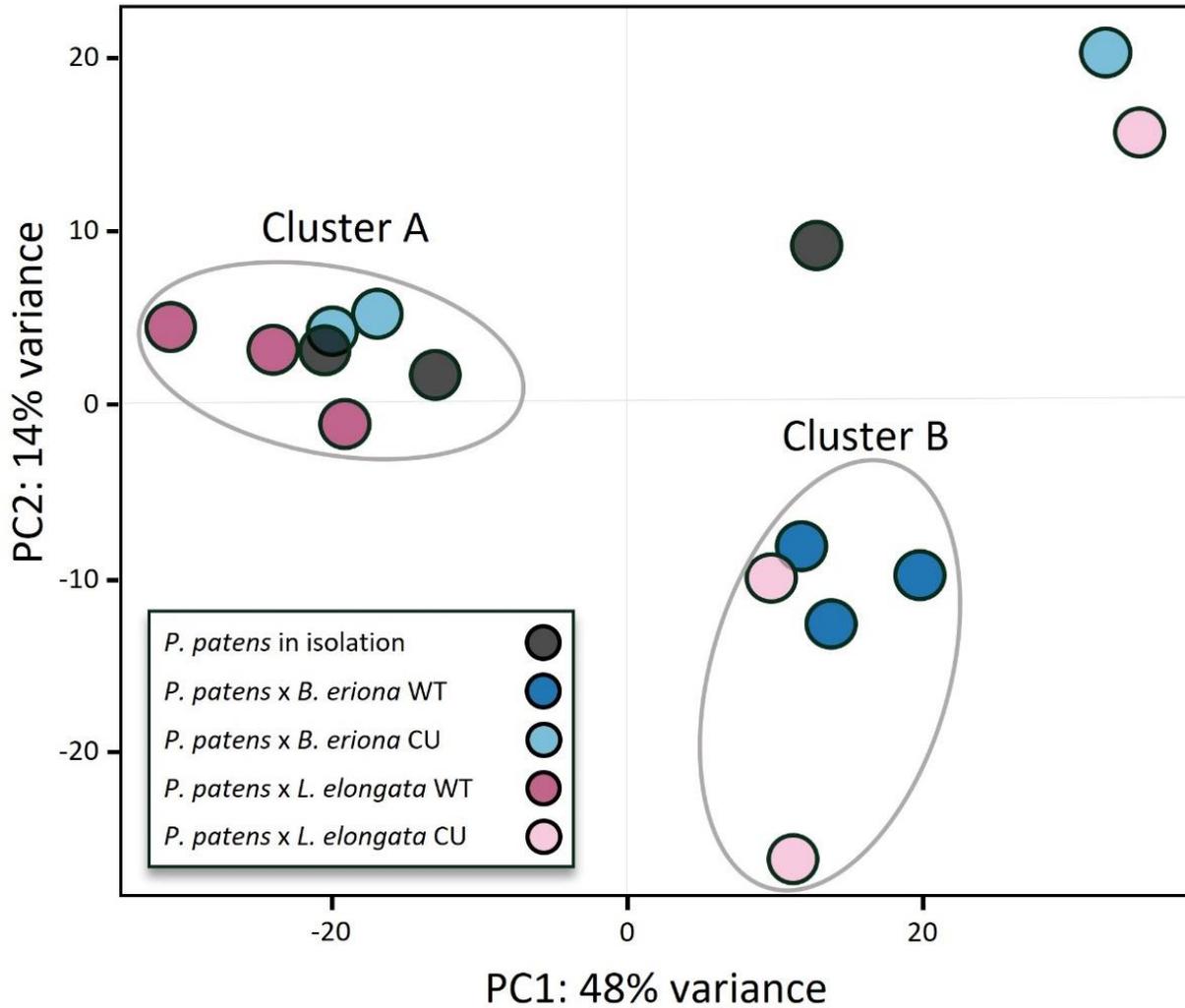
Our study consisted of 15 transcript libraries, containing 5 different treatment groups with 3 replicates each. Among the metadata, mapped reads mainly aligned to the *P. patens* transcriptome ( $99.1 \pm 1.2\%$  (95% CI)). The 0.9% of reads to map to a fungal transcriptome were

heavily inflated by replicates in *P. patens* x *L. elongata* CU cocultures (CUTPG, CUTPC, and CUTPH), which represented 0.5%, 1.7% and 8.4% of total mapped reads respectively. This is in contrast to the other 9 fungal coculture samples, which generally mapped less than 0.01%. The pool of total mapped fungal reads, even with *L. elongata* CU samples, was not sufficiently representative for analysis of transcriptomic response in fungi. Our principal component analysis (PCA) showed that the derived variance between samples led to two major clusters based on transcriptomic response (Figure 2.3). Consistent with the distinct microscopic interaction observed, cocultures indicating stress responses in *P. patens* (Cluster B: *B. erionia* WT and *L. elongata* CU) clustered separately from samples speculated to induce a neutral response (Cluster A: uninoculated *P. patens*, *B. erionia* CU, and *L. elongata* WT (Figure 2.3)).

Additionally, investigating the identified gene homologs necessary for plant-fungal symbiosis [Delaux et al. 2015] presents strong differences between WT and CU strains of each fungal species effects on *P. patens* (Supplemental Data 7). The presence of endobacteria seems to influence the *P. patens* symbiotic genes in these strains inversely, where *B. erionia* CU and *L. elongata* WT have no hits and only one differentially expressed gene (DEG; a downregulated GRAS transcription factor) respectively represented in both strains. That contrasts strongly with *B. erionia* WT and *L. elongata* CU with 20 DEGs and 14 DEGs respectively and among those hits, 8 DEGs (1 MLD-Kinase; 2 CDPKs; 5 GRAS transcription factors) were shared and regulated the same between both strains (Supplemental Data 7).

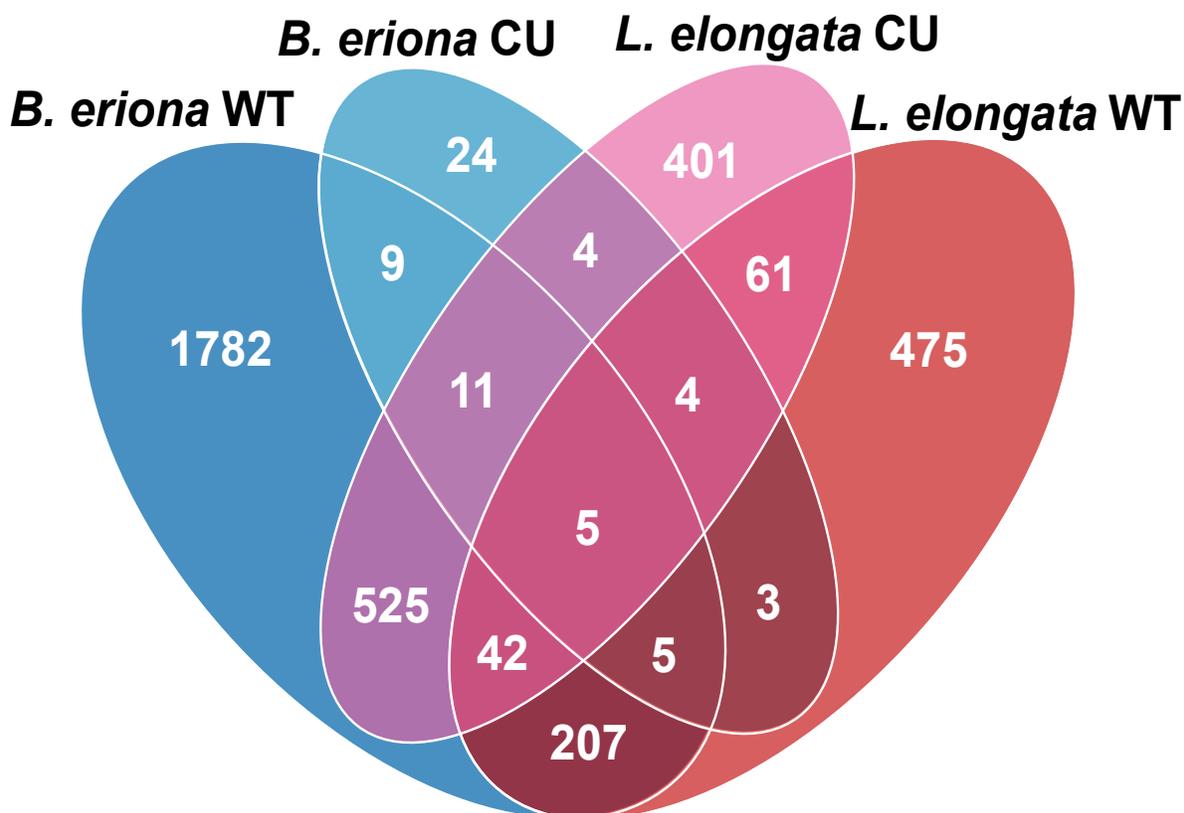
Of all the treatments investigated (fungal presence, *B. erionia* coculture, *L. elongata* coculture, and endobacteria presence/absence) we identified that the combined effects from species and endobacteria were the most informative as each induced unique responses in *P. patens*. This phenomenon is especially illustrated by 75.4% of all DEGs being unique to a specific fungal

coculture (Figure 2.4; Supplemental Data 3a-d).



**Figure 2.3:** Principal Component Analysis (PCA) of *P. patens* mapped RNA-seq reads for the 15 RNA-sequencing libraries generated with DeSEQ2. The color of each point correlates to experimental treatment.

*P. patens* control grown in isolation (black), and *P. patens* treatments grown in coculture with *B. erionia* WT (dark blue), *B. erionia* CU (light blue), *L. elongata* WT (magenta), and *L. elongata* CU (pink).



**Figure 2.4:** Venn diagram of DEGs between *P. patens* control and *P. patens* co-cultures with *B. erionia* WT (dark blue), *B. erionia* CU (light blue), *L. elongata* WT (red), and *L. elongata* CU (pink).

Total DEGs called by DESeq2 and their significance between each treatment. Genes qualified as differentially expressed if they had expression differences with  $P_{adj} < 0.01$  compared to *P. patens* control.

***Transcriptomic response from B. erionia WT suggest infectious activity in P. patens***

Among the four fungal treatments, *P. patens* cocultured with *B. erionia* WT had the most definitive response phenotypically and transcriptomically. There were 2,586 total significant DEGs ( $P_{adj} < 0.01$ ), with 1,533 upregulated and 1,053 downregulated (Supplemental Data 3b). Gene groups with overrepresented expression in *B. erionia* WT cocultures had ontology enrichment especially with function in transmembrane transport, calcium/inorganic ion transport, and localization (Supplemental Data 9). Consistent with the observed chloroplast rerouting

between colonized plant cells, the enrichment of organelle transport genes supports that *P. patens* transcriptomically responds to contain fungal infection and limit organelle damage by changing cellular organization, which is an observed infection response [Savage et al. 2021]. Of specific transporters, a putative syntaxin transporter was highly represented among *B. erionia* WT coculture gene hits and is shown to function in cellular reorganization for the shuttling/protection of organelles in other systems (Table 1) [Hachez et al. 2014]. Ion transport, particularly calcium, was also overrepresented. Calcium channels are essential for cell signaling locally and globally within plant systems, notably influencing immune response for both parasitic and symbiotic fungi [Chen et al. 2015A, Ivashuta et al. 2005] and also can direct the tethering or relocation of organelles [Allan et al. 2022, Tominaga et al. 2012]. Additional transporters from DEGs included a chloroquine resistance transporter, whose homologs have been affiliated with regulating abiotic stress in *Arabidopsis* [Maughan et al. 2010, Waller et al. 2003], as well as a nodulin-like transporter. Nodulin-like transporter homologs have been observed in other non-nodulating systems and play important roles not only in the transport of micronutrients but also are critical for communication between the plant-fungal interface for symbiosis and infection [Akiyama et al., 2005, Besserer et al., 2006, Denancé et al. 2014, Waters et al., 2013].

Pathways influencing photosynthesis, abiotic stimuli response, and transcriptional regulation saw substantial depletion in expression (Supplemental Data 9). Photosynthesis changed most prominently with 28 related GO terms having decreased expression in coculture, including chlorophyll binding, photosystem I & II, thylakoid activity, and chloroplast activity (Supplemental Data 9). Photosynthesis pathways play an important role in plant immunity. Reduced photosynthesis activity has been reported as one of the first actions in immune response to both abiotic and biotic stressors [Lu & Yao 2018, Yang et al. 2022]. Typically, during the early

stages of infection photosynthetic activity is reduced and even after infection has passed transcript accumulation can remain low [Chen et al. 2015B, Hu et al. 2020, Scharte et al. 2005, Swarbrick et al. 2006, Yang et al. 2022].

We also investigated highly expressed individual genes outside the gene ontology categories, whose differential expression posed interesting considerations. Carbon metabolism genes stood out due to their distinct role in mediating plants-fungal symbiosis [Bonfante & Genre 2010].

Many of the highest supported and upregulated genes presented here are annotated with putative functionality in carbohydrate synthesis and cellulose/cell wall biosynthesis (Table 1).

Additionally, there was also activation of three transcription factors which may be involved in the regulation of the previously highlighted gene ontologies. These included a WRKY, TIFY, and HEX transcription factor (Table 1). WRKY transcription factors are heavily represented and conserved among embryophytes, with nearly 40 copies in *P. patens*, and regulate the expression of abiotic stress, biotic stress, and developmental response [Bakshi & Oelmüller 2014]. In angiosperms, TIFY motif transcription factors have demonstrated the signaling of growth, development and defense response, and we suspect similar activity here [Xia et al. 2017]. *B. erionia* WT cocultures identified upregulation of a hydroxymethylglutaryl-CoA reductase (HMG-CoA; Table 1), encoding a key regulatory enzyme in the cytosolic mevalonate (MVA) terpenoid biosynthetic pathway (Table 1) [Friesen & Rodwell 2004, Simkin et al. 2011]. This gene is involved in the production of precursors of sterol biosynthesis, which are integral for membrane integrity and hormonal responses [Morikawa et al. 2009]. The transcriptomic and histological responses described here parallel those recently described in the infection of *P. patens* with the broad spectrum necrotrophic fungal pathogen *Botrytis cinerea* (Sclerotiniaceae) and provides

strong support for classifying *B. erionia* WT as pathogenic towards *P. patens* [Reboledo et al. 2021, Reboledo et al. 2020].

***When cured of endobacteria, B. erionia loses capacity for interaction with P. patens***

In contrast to the response with *B. erionia* WT, the strain cured of the endobacterium reduced transcriptional response 40-fold when cocultured with *P. patens* (Figure 2.4). We detected 65 DEGs ( $P_{\text{adj}} < 0.01$ ) in the presence of *B. erionia* CU, with only 46 upregulated genes and 19 downregulated genes (Supplemental Data 3c). This limited set of genes still shows ontology enrichment in pathways for extracellular sensing and cell wall response (Supplemental Data 9). There were no gene ontologies depreciated in *P. patens* when in coculture with *B. erionia* CU. We identified two upregulated transcription factors, which included an ethylene responsive transcription factor and a MYB-like transcription receptor. Ethylene responsive transcription factors are conserved throughout embryophytes and regulate many diverse regulatory pathways but predominantly are involved in response to external stimuli [Binder 2020, Hall et al. 1977, Licausi et al. 2013]. Notably, the key committed step involved in *ent*-kaurene biosynthesis, encoded by the *ent*-kaurene synthase gene (Table 1), was also identified to be downregulated here.

*Ent*-kaurene has parallel functions to gibberellins in angiosperms, which generally function in the regulation of developmental changes and response to pathogens. The downregulation of *ent*-kaurene synthase here provides implication for decreased environmental sensitivity [Hayashi et al. 2010, Miyazaki et al. 2011, Miyazaki et al. 2018, Reboledo et al 2021, Reboledo et al. 2020]. Overall, *P. patens* response to *B. erionia* CU presents itself as relatively neutral, non-specific, and with no characteristics of plant-pathogen interaction. Therefore, *B. erionia* appears to require the endobacterium to retain pathogenicity.

### ***L. elongata* WT cocultures demonstrate beneficial tendencies with *P. patens***

When *L. elongata* WT was cocultured with *P. patens*, a total of 802 genes were differentially expressed ( $P_{\text{adj}} < 0.01$ ), with 287 genes upregulated and 515 genes downregulated (Supplemental Data 9). Of the upregulated genes, we identified enrichment in three ontologies: cytoskeletal reorganization, cell wall biogenesis and the synthesis of various polysaccharide pathways (Supplemental Data 9). Complementing what was observed with microscopy, cytoskeletal rearrangement in *P. patens* may be occurring to harbor the *L. elongata* hyphae within plant cells but the cytoskeleton also plays dynamic roles in plant growth, development, and immune response [Wang et al. 2022]. Additionally, we identified enrichment in cell wall biosynthesis (Supplemental Data 9). Differentiation of cell wall activity likely has a related function to the differentiation of cytoskeletal rearrangement since both are implicated in cellular architecture. Upregulation for the synthesis of polysaccharides may be relevant in two different pathways due to the integral role of polysaccharides in the cell wall [Voiniciuc et al. 2018] or due to the essential role carbon exchange plays in plant-fungal mutualism [Bonfante & Genre 2010]. Alterations to carbon metabolism are generally indicative of plants-fungal symbiosis and the higher expression for these pathways here may suggest a positive interaction between these two species. Notable DEGs directly involved in carbon/photosynthesis included the Chlorophyll A/B binding protein, beta-1-3 glucanase, and GDP-fucose transferase (Table 1). These genes also suggest a heightened production of polysaccharides and also may indicate the upregulation of carbon metabolism. Further support for cross kingdom interaction comes from DEGs encoding a formerly characterized alpha- dioxxygenase (Table 1), which has been shown to participate in fungal infection response and plant development [Groenewald and Weshuizen 1997, Machado et

al. 2015], as well as a spermidine synthase (Table 1), which has implications in plant host defense against infection [Mueller 1998, Stenzel et al. 2003, Takahashi & Kakehi 2010]. Comparatively, *L. elongata* WT uniquely caused a disproportionately high number of downregulated genes including a depleted response to oxidative stresses, metal ion binding, and photosynthesis (Supplemental Data 9). Reduced sensitivity to oxidative stresses is a common response induced by endophytic fungi in plant systems [Clay 1988, Fontana et al. 2021, White and Torres 2010]. This relationship often embodies a mutualistic interaction by aiding both systems in defense, where fungi provide a heightened protection against abiotic stressors, specifically reactive oxygen species (ROS) [Clay 1988, Fontana et al. 2021, White and Torres 2010]. This can be accompanied by fungal ROS mediation of the host, creating “leaky” plant cells enabling easier access to nutrients by fungal endophytes, and could further help explain the previously highlighted DEG, pectate lyase (Table 1) [Su 2023, White and Torres 2010]. In plants, iron plays a key role in photosynthesis and the repression of its ontology here coincides with reduced photosynthetic activity (Supplemental Data 9). Endophytic fungi have been reported to provide absorbable iron from the soil to their plant hosts [Verma et al. 2022]. This may indicate that *P. patens* is receiving iron from *L. elongata* WT and consequently reducing iron binding, consistent with the observed repression of a ferritin like receptor (Table 1). Like with *B. erionia* WT, we also detected a strong depletion of gene expression affiliated with photosynthesis in *P. patens* x *L. elongata* WT cocultures. This repression is likely again due to the important role of photosynthesis in general immune response. A distinction for *L. elongata* WT photosynthetic response compared to *B. erionia* WT, is the affiliated deactivation of many early light-induced proteins (Table 1), which are usually activated in response to abiotic stress [Hutin et al. 2003]. This complex suggests that an induced immune response is occurring, which, in combination

with the lack of any asymptomatic phenotypes to photosynthetic tissue, may further point to the establishment of a beneficial interaction in *P. patens* x *L. elongata* WT cocultures.

***Endobacterial absence in L. elongata CU influences its environmental colonization and shifts its subsequent interaction with P. patens to resemble B. erionia WT response***

*P. patens* in coculture with *L. elongata* CU had a total of 1053 DEGs ( $P_{\text{adj}} < 0.01$ ), with 829 genes upregulated and 224 genes downregulated. While the quantity of DEGs was comparable to *L. elongata* WT, only 112 (10.3%) DEGs were shared between both *L. elongata* strains. In contrast 583 (53.8%) DEGs were shared with *B. erionia* WT (Figure 2.4). We identified enrichment pathways in cell periphery/membrane activity, transcription factor activity, and lipid transport (Supplemental Data 9). As with *B. erionia* CU, *L. elongata* CU also displayed no significant depletion of specific gene ontologies (Supplemental Data 9). Due to the characteristically high density of chlamydospores from *L. elongata* CU cocultures and enrichment of cell periphery, the changes in *P. patens* gene expression may suggest mechanical or chemical interactions distinct from the other cocultures due to the aforementioned chlamydospores. As in *B. erionia* CU, the activation of the same ethylene response transcription factor (Table 1) could implicate signaling of pathogen response and/or alternative cell development in *P. patens* [Binder 2020, Hall et al. 1977, Licausi et al. 2013]. The upregulation of phosphorelay signal transduction system and hybrid signal transduction histidine kinase (Table 1), involved in the regulation of osmotic and oxidative stress, may indicate that *L. elongata* CU is inducing a stress response cascade in *P. patens* [Carapia-Minero et al. 2018]. Many of the differentially expressed transcription factors identified here have homologs which are directly involved in other systems for growth, development, signaling, and differentiation. These included a HEX motif transcription factor [Soufi & Jayaraman 2008], an E2F/DP family helix DNA binding protein [Müller et al. 2001, Mariconti et al. 2002], a BIM1 motif transcription factor specifically induced through membrane

signaling [Yin et al. 2004], an RNA Pol II transcription regulator co-expressed with sporophyte development, and a cycling DOF factor [Ishida et al. 2014, Goralogia et al. 2017, Wei et al. 2018] (Table 1). We identified an allene oxide synthase (Table 1) with suggested role in the jasmonate pathway and potential involvement in development and stress response [Stenzel et al. 2003]. Also, we identified the same HMG-CoA reductase differentially expressed in *B. erionia* WT, indicating a connection to plant defense [Friesen & Rodwell 2004 Simkin et al. 2011]. Genes with lower transcript accumulation in *L. elongata* CU cocultures included a GLK1 motif transcription factor (Table 1) which has conserved function throughout embryophytes (and the algae *C. reinhardtii*), directly influencing chlorophyll biosynthesis [Gang et al. 2019, Waters et al. 2009, Yasumura et al. 2005]. Represented as well were genes encoding a putative xyloglucan glycosyltransferase and an alpha-ketoglutarate sulfonate dioxygenase, which are involved in carbon metabolism and more specifically saccharide production. The upregulation of these genes could suggest either that carbon exchange may be occurring or, because the plant cell wall is a major sink for saccharides, these genes could also indicate major changes to cell wall structure instead.

**Table 2.1:** Select ontology and gene representatives from differential gene expression among *P. patens* and Mortierellaceae cocultures

Representative differentially expressed genes from each coculture that exemplify the enriched/depleted ontologies and the unique interactions. For selected genes with particularly strong support and with implication in plant-fungal exchange but with ontologies not significantly represented (n.s.) were also included in this dataset.

Coculture	Ontology	Regulation	Gene(s)	Expression	Annotation	Log2	Padj
<i>B. eriona</i> WT	transport / localization	Enriched	Pp3c19_4690	↑	syntaxin transporter	4.23	1.14x10 <sup>-40</sup>
		Enriched	Pp3c16_7270	↑	membrane protein	2.12	7.01x10 <sup>-40</sup>
		Enriched	Pp3c14_6070	↑	ATP-binding cassette transporter	2.28	3.12x10 <sup>-31</sup>
		Enriched	Pp3c1_4260	↑	predicted transporter protein	1.77	2.60x10 <sup>-45</sup>
		Enriched	Pp3c9_2960	↑	nodulin like transporter	3.08	6.06x10 <sup>-28</sup>
	carbon metabolism	Enriched	Pp3c6_18950	↑	chloroquine resistance transporter	2.69	6.06x10 <sup>-28</sup>
		n.s.	Pp3c4_25090	↑	carboxykinase	2.77	4.00x10 <sup>-40</sup>
		n.s.	Pp3c14_11970	↑	exostosin related gene	2.07	5.74x10 <sup>-36</sup>
		n.s.	Pp3c21_16620	↑	cellulose/cell wall biosynthesis	3.38	2.47x10 <sup>-36</sup>
		n.s.	Pp3c5_19530	↑	cellulose/cell wall biosynthesis	1.96	3.17x10 <sup>-26</sup>
		n.s.	Pp3c1_41400	↑	cellulose/cell wall biosynthesis	1.72	3.21x10 <sup>-22</sup>
	transcription factor	n.s.	Pp3c15_25660	↑	cellulose/cell wall biosynthesis	1.68	5.85x10 <sup>-22</sup>
		n.s.	Pp3c19_3000	↑	WRKY transcription factor	3.31	3.12x10 <sup>-31</sup>
		n.s.	Pp3c19_8700	↑	protein TIFY	2.36	3.86x10 <sup>-31</sup>
	isoprenoid synthesis	n.s.	Pp3c6_2730	↑	HEX transcription factor (leaflet specific)	2.17	4.12x10 <sup>-23</sup>
n.s.		Pp3c1_10000	↑	hydroxymethylglutaryl-CoA reductase	1.57	2.94x10 <sup>-30</sup>	
stress response	n.s.	Pp3c19_17670	↓	BAG family molecular chaperone regulator	-2.22	3.19x10 <sup>-27</sup>	
	n.s.	Pp3c19_10440	↓	DNAJ homolog subfamily	-1.54	1.04x10 <sup>-25</sup>	
<i>B. eriona</i> CU	extracellular sensing	Enriched	Pp3c26_13260	↑	ethylene response transcription factor	1.68	2.80x10 <sup>-4</sup>
	cell wall response	Enriched	Pp3c23_11030	↑	MYB-like transcription factor	1.83	4.01x10 <sup>-4</sup>
		Enriched	Pp3c4_24020	↑	cell wall assembly regulator	1.51	9.86x10 <sup>-6</sup>
	isoprenoid synthesis	n.s.	Pp3c7_1880	↓	ent-kaurene synthase	-1.50	1.91x10 <sup>-6</sup>
<i>L. elongata</i> WT	cytoskeletal reorganizaiton	Enriched	Pp3c18_1430	↑	myosin ATPas	0.90	2.19x10 <sup>-6</sup>
	cell wall biogenesis / polysaccharide synthesis	Enriched	Pp3c5_7180	↑	chlorophyl A/B binding protein	1.29	3.07x10 <sup>-10</sup>
		Enriched	Pp3c17_5150	↑	beta-1-3-glucanase	1.45	1.23x10 <sup>-7</sup>
		Enriched	Pp3c3_20980	↑	GDP-fucose transferase	0.90	5.11x10 <sup>-6</sup>
		Enriched	Pp3c17_16370	↑	pectate lyase	0.98	6.45x10 <sup>-8</sup>
	immune response	n.s.	Pp3c26_4220	↑	alpha-dioxygenase	1.57	4.29x10 <sup>-6</sup>
		n.s.	Pp3c6_27380	↑	spermidine synthase	1.00	4.89x10 <sup>-6</sup>
	metal ion binding	Depleted	Pp3c7_6750	↓	ferritin like receptor	-1.30	1.22x10 <sup>-14</sup>
	photosynthesis	Depleted	Pp3c24_9670	↓	early light induced proteins	-2.07	1.48x10 <sup>-22</sup>
Depleted		Pp3c11_7280	↓	early light induced proteins	-1.89	1.16x10 <sup>-12</sup>	
<i>L. elongata</i> CU	transcription factor	Enriched	Pp3c26_13260	↑	ethylene response transcription factor	2.25	7.28x10 <sup>-11</sup>
		Enriched	Pp3c6_2730	↑	HEX motif transcription factor	1.93	1.49x10 <sup>-10</sup>
		Enriched	Pp3c22_10160	↑	E2F/DP family helix DNA binding protein	2.95	2.18x10 <sup>-10</sup>
		Enriched	Pp3c9_470	↑	BIM1 motif transcription factor	1.82	6.92x10 <sup>-10</sup>
		Enriched	Pp3c11_6620	↑	RNA pol II transcription regulator	2.33	6.92x10 <sup>-10</sup>
	lipid transport	Enriched	Pp3c17_3860	↑	cyclic DOF factor	2.57	6.92x10 <sup>-10</sup>
		Enriched	Pp3c1_10000	↑	hydroxymethylglutaryl-CoA reductase	1.40	9.05x10 <sup>-11</sup>
	carbon metabolism	Enriched	Pp3c5_3730	↑	allene oxide synthase	3.02	3.66x10 <sup>-11</sup>
		n.s.	Pp3c21_16620	↑	xyloglucan glycosyltransferase	3.19	9.39x10 <sup>-13</sup>
	photosynthesis	n.s.	Pp3c20_13320	↑	alpha-ketoglutarate sulfonate dioxygenase	1.60	3.28x10 <sup>-11</sup>
	n.s.	Pp3c7_5800	↓	GLK1 motif transcription factor	-2.69	1.25x10 <sup>-13</sup>	

### Comparisons with the ‘Gene Atlas Project’ identifies multiple novelties in *P. patens* expression

‘The *Physcomitrella patens* gene atlas project: large-scale RNA-Seq based expression data’

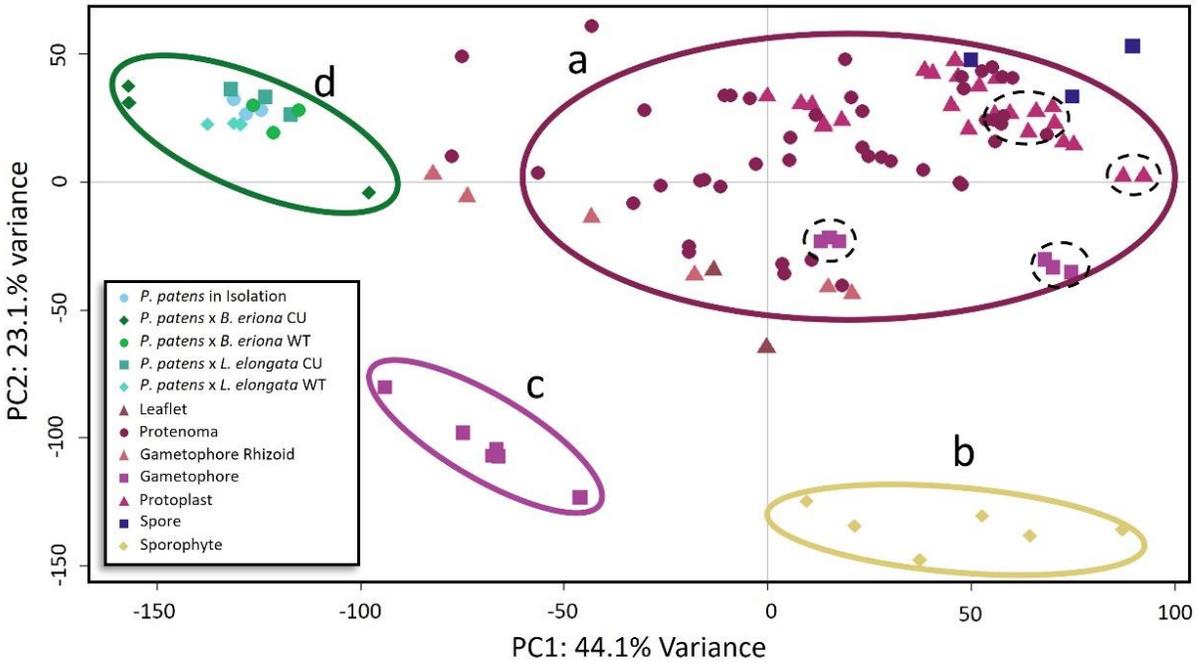
(‘Gene Atlas Project’) was published in response to the updated *P. patens* chromosome-scale

genome assembly V3.3 in 2018 [Perroud et al. 2018, Lang et al. 2018]. This resource was developed to identify specific trends involved in developmental stages, environmental conditions, and to support the reproducibility of RNA-sequencing in *P. patens* amongst different labs [Perroud et al. 2018]. In contrast to the samples of the ‘Gene Atlas Project’, grown aseptically on defined growth media, we introduced two novel variables, fungal cocultures and growth on soil, a substrate closer resembling nature with no noticeable impact on *P. patens* or *Mortierella* growth. Comparisons between all samples (our 15 and the 99 ‘Gene Atlas Project’ samples) were investigated using a principal component analysis (PCA) (Figure 2.5) (Supplemental Code 4). Our samples (d) clustered tightly and distinctly from the original three major clusters “gametophores” (a), “sporophytes” (b), and “strong stresses” (c). With PC1 explaining 44.1% of variance, PC2 explaining 23.1% and PC3 explaining 11.1%, the inclusion of *P. patens* grown with fungi and on soil sufficiently shifted the original ‘Gene Atlas Project’ dispersion while maintaining the same clustering patterns from the originally explained variation (PC1:78%, PC2:10%, PC3:6%) [Perroud et al. 2018]. Growth on soil had a more defining signal than fungal cocultures, seen with the tight distribution of all 15 samples, including the control replicates. Notably, the dotted circles in Figure 2.5 represent control experiments grown on the conventional liquid/solid BCD and Knop medias with the same *P. patens* strain, genotype and tissue types sampled as those grown here on soil [Cove et al. 2009, Perroud et al. 2018, Reski & Abel 1985], which especially highlights the influence growth media has on gene expression. Explanatory variables for the expression difference of soil compared to lab-based agar/liquid medias include supplements such as CaCO<sub>3</sub> and MgCO<sub>3</sub> rich dolomite, silicon, and organics through peat moss. Additionally, the soil provides moisture retention and structural aeration different than agar-based medias. Although all samples tested here differed substantially from

one another, growth on soil was the most influential factor in distinguishing their expression from the ‘Gene Atlas Project’ (Figure 2.5). The impact media has on *P. patens* expression is significant as “control” samples grown on common *P. patens* growth medias including BCD Agar, BCDAT, Knop Liquid/Agar, Hoagland, and PpNH<sub>4</sub> protoplast medias all lead to distinct clustering of samples despite each recipe, other than soil, having similar composition (Supplemental Data 5). Additionally, soil samples are on the same axis with PpNH<sub>4</sub> and Hoagland based media for PC1 (70.9% variance) and with Knop/BCD Agar samples for PC2 (19.6% variance) (Supplemental Data 5). While using soil as a media here also leads to distinct clustering of samples from other medias, they do not vary substantially to warrant their own category entirely. Comparatively, we see the difference in expression from growth on soil or growth on Knop agar is comparable to the difference from growth on BCD agar or Knop agar (Supplemental Data 5). Differential expression analysis comparing our samples and the ‘Gene Atlas Project’ produced a total of 7,450 significant DEGs ( $P_{adj} < 0.0001$ ). Among these, 3,873 were upregulated in our samples and 3,577 were downregulated (Supplemental Data 4a). Among upregulated genes, we identified enrichment in ubiquitin expression, transcription and translation, and nitrogen biosynthesis pathways (organonitrogen, peptides, and amides) (Supplemental Data 9). From downregulated genes, we identified depleted activity in mitosis and cell division, endoplasmic reticulum activity, and Golgi apparatus activity (Supplemental Data 9). A total of 2,116 genes lacked any sign of expression specifically in our samples but were expressed in at least one other ‘Gene Atlas Project’ sample (Supplemental Data 4e). Genes silenced in our system but expressed elsewhere saw enrichment in membrane processes and DNA transposition (Supplemental Data 9). This included 7 genes with expression conserved in all 99 ‘Gene Atlas Project’ samples and silenced in our system (Supplemental data 4e). Two of

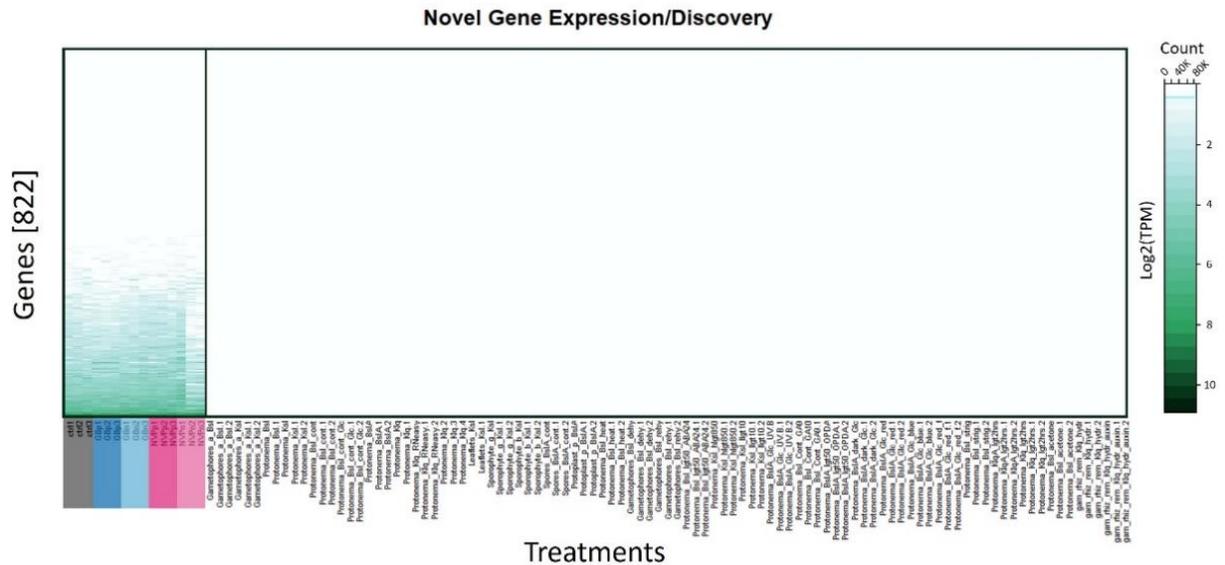
these genes function in endoplasmic reticulum transport (Pp3c9\_10380, Pp3c12\_8160) and three were involved in regulation of the central dogma (Pp3c1\_24700, Pp3c11\_25450, Pp3c10\_6100). In contrast, novel expression exclusive to the current study compared to the ‘Gene Atlas Project’ yielded a total of 822 genes (Figure 2.6). Among those, 55 genes (6.7%) were not scaffolded to a chromosome, representing over a 3-fold higher abundance than what would be expected based on the transcriptomic makeup. Additionally, from those 822 total genes, 26% had significantly high differential expression with the most highly supported gene (Pp3c15\_19571) having an average of 781 TPM (in the top 1% of genes expressed). Among genes with novel expression, we detected enrichment in pathways involved in RNA polymerase II activity, cell shape regulation, and beta (1,3)-D-glucan biosynthesis activity (Supplemental Data 9). While RNA polymerase II activity is most well-known for its role in the production of mRNA, it also plays a dynamic and conserved role in plant pathogen response and its representation here likely shares both functionalities [Li et al. 2014]. A sub-class of RNA-polymerase II, the RNA-directed RNA polymerases, also has utility in forming small interfering RNA (siRNA) [Du et al. 2022, Hunter et al. 2013] and a putative RNA-directed RNA polymerase was included among our identified genes (Pp3c6\_8521). Generally, siRNAs function against the invasion of foreign entities, alternative gene regulation, and control of transpositional elements [Castel & Martienssen 2013, Du et al. 2022], which may parallel the unique challenges presented in our conditions. Also, DNA-transposition was one of the few ontologies underrepresented in our system and its reduced transposition may be due to the presence of siRNA. Enrichment of cell shape morphology detected in all samples may imply the necessity for certain cellular features for soil specific conditions. Genes putatively involved in beta (1,3)-D-glucan biosynthesis have a presumed

involvement in cell wall structure [Douglas 2001, Roberts et al. 2012, Roberts et al. 2018] and plant defense. [Vega-Sánchez et al. 2013].



**Figure 2.5:** Principal component analysis (PCA) of transcripts per million mapped reads (TPM) of the 99 libraries from the ‘Gene Atlas Project’ [Perroud et al. 2018] and the 15 samples analyzed here.

Samples had four distinct clusters with various gametophore tissues (maroon, a), sexual tissues (yellow, b), high heat samples (purple, c), and samples grown on soil (green, d). Circled with dotted ellipses are control groups from intra- and interlaboratory conditions with different medias (Knop Agar, Knop liquid, BCD Agar). PC1 explains 44.1% of the variance and PC2 explains 23.1% of the variance.



**Figure 2.6:** Heatmap of the 822 genes with mapped reads to the dataset were presented here and absent among all ‘Gene Atlas Project’ samples.

Rows ordered based on the total sum of gene expression per gene. Columns represent the 114 sample conditions. Control groups (black), *B. erionia* WT (dark blue), *B. erionia* CU (light blue), *L. elongata* WT (magenta), *L. elongata* CU (pink), and ‘Gene Atlas Project’ samples (white). Expression in TPM is measured based on a log<sub>2</sub> scale where the darker shade of green indicates more expression and white indicates no expression.

### ***Comparative Expression Across P. patens, A. thaliana, and C. reinhardtii identifies Shared Trends***

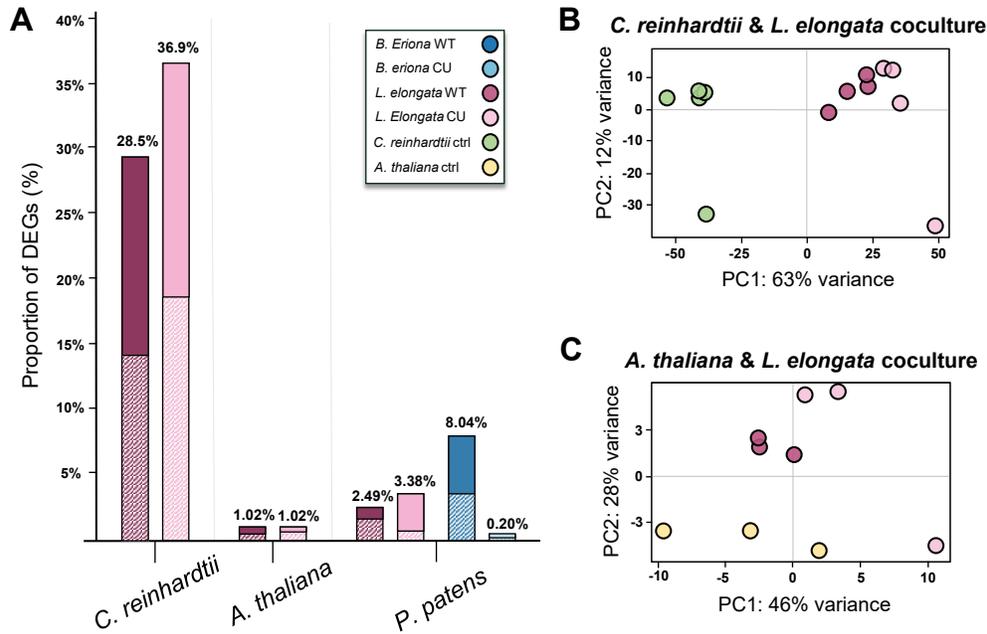
To determine potentially conserved traits between taxonomically distant plant lineages, the responses from *P. patens* in coculture with *L. elongata* WT and CU were compared to *Arabidopsis thaliana* (*A. thaliana*) and *Chlamydomonas reinhardtii* (*C. reinhardtii*) [Vandepol 2022]. Comparatively, the response in *A. thaliana* only represented a smaller subset of the transcriptome where 280 and 279 genes were differentially expressed with *L. elongata* WT and *L. elongata* CU cocultures, respectively (Figure 2.7A). Despite lower overall DEGs, improved aerial growth was still observed in *L. elongata* coculture regardless of endobacterial presence [Vandepol 2022]. Remarkably, and consistent with the interaction being inherently mutualistic, only *C. reinhardtii* in coculture survived two weeks of growth in minimal conditions, as both the

algae and fungi in isolation were unable to do so. Due to the harsh conditions, significant changes to gene expression were observed, with 4,363 and 5,636 genes differentially expressed in *L. elongata* WT and *L. elongata* CU cocultures, respectively (Figure 2.7A). Additionally, for both *A. thaliana* and *C. reinhardtii*, the principal component analysis showed clustering of samples mainly based on coculturing and less influence from endobacteria presence/absence than what was observed with *P. patens* (Figure 2.7B & 2.7C).

Based on phenotypes and transcript profiles *L. elongata* WT appeared to establish a beneficial response among cocultures with all plant hosts analyzed here. From shared enrichment in ontology, the transcriptomic response induced *in plantae* by *L. elongata* WT also shared similarities despite the vast degree of divergence from *P. patens* (~630 MYA for *C. reinhardtii*, and ~450 MYA for *A. thaliana*). *P. patens* cocultures had influence on large scale cellular reconfiguration, increased carbon metabolism, reduced susceptibility to abiotic stress, and reduced photosynthesis. Likewise, *C. reinhardtii* also displayed an increased representation of carbohydrate metabolism largely in the form of glycosylation activity, reduced photosynthesis, and reduced response to abiotic stress (Supplemental Data 4). *A. thaliana* also showed an increase of carbon metabolic processes but inversely increased photosynthetic activity and stress response (Supplemental Data 9). A shared differential expression of isoprenoid/terpene metabolism, observed among all three plants investigated here, could represent a defining and conserved feature across species. Unique to *P. patens* cocultures, the presence and absence of endobacteria distinctly influenced the interaction *in plantae*.

We investigated the response of single copy orthologs between *P. patens*, *A. thaliana*, and *C. reinhardtii* (Supplemental Data 8c-g). All three species shared a total of 1,524 single copy orthologs, *P. patens* and *A. thaliana* shared 3,032 single copy orthologs, and *P. patens* and *C.*

*reinhardtii* shared 2,484 single copy orthologs. Only one single copy ortholog shared differential expression patterns between the plant hosts (Pp3c9\_570; AT4G24160; CHLRE\_01g000300) and is characterized in *A. thaliana* for maintaining the lipid homeostasis by regulating both phospholipid and neutral lipid levels. Generally, this gene plays an important role in regulating glycerol-3-phosphate availability and downstream lipid metabolism [Brzoska & Boos 1988, Larsen et al. 1983, Raetz 1986, Wang et al. 2016]. *P. patens* and *C. reinhardtii* shared a total of 115 additional single copy orthologs (Supplemental Data 8c). Among those, there are 20 orthologous pairs coregulated in coculture with *L. elongata* WT and/or *L. elongata* CU. These orthologs may represent parallel expression patterns conserved across 600 MY of divergence and may hold key insights to the long-standing, fundamental mechanisms of plant-fungal symbiosis.



**Figure 2.7:** Mortierella coculture DEGs ( $P_{adj} < 0.01$ , DESeq2) comparisons across *C. reinhardtii*, *A. thaliana*, and *P. patens*.

A.) Proportion of significant DEGs compared within each respective transcriptome (*C. reinhardtii*: 17,469 genes; *A. thaliana*: 27,459 genes; *P. patens*: 32,161 genes) among cocultures with *B. erionia* WT (blue), *B. erionia* CU (light blue), *L. elongata* WT (magenta), *L. elongata* CU (pink). Solid fill, upregulated genes in each sample; striped sections, downregulated genes

B.) PCA distribution of expression of *C. reinhardtii* in isolation (green), coculture with *L. elongata* WT (magenta) or *L. elongata* CU (pink).

C.) PCA distribution of *A. thaliana* in isolation (yellow), coculture with *L. elongata* WT (magenta) or *L. elongata* CU (pink).

## Conclusion

Here we report that the Mortierellaceae species *B. erionia* and *L. elongata* are both capable of endophytic interactions with *P. patens*, albeit subtle and possibly dependent on endobacteria. The ability for these fungi to interact with *P. patens* is largely dependent on endobacterial presence and absence, which changes both the plant phenotypic and transcriptomic response. Current understandings of how endobacteria influence their fungal host and the subsequent effects that these have on their environment are still limited, and because of this, their influence on *P. patens* interaction is notable, particularly with MRE in *B. erionia*. Interactions between any two

organisms are environmentally dependent and may be beneficial or adversarial depending on those conditions, [Bonfante & Genre 2010, Dickie et al. 2013, Du et al. 2019, Eastburn et al. 2011, Giaouque & Hawkes 2013] and as such we are just beginning to uncover the complexities of potential relationships with *P. patens* and fungi in Mortierellaceae.

Comparison with the ‘Gene Atlas Project’ indicated distinct changes to gene expression caused by different growth conditions (soil vs. common lab-based medias). Comparisons across our work and that of the ‘Gene Atlas Project’ yielded the discovery of 822 genes with novel expression, 7 genes which previously were otherwise constitutively on, and an updated “Gene Atlas” reference with the appending of our data along with methodology to build upon these methods moving forward (Supplemental Data 4a-b; Supplemental Code 4). While the mechanisms of how plants initially colonized land remains a mystery, the biodiversity and response captured by the second largest clade of land plants, the bryophytes, helps to provide further gravity to the influential role fungi played in making that possible. In conclusion, the exchange between plant hosts and fungal symbionts, and the evolution of those exchanges, are dynamic, competitive, and conditional.

### **Data availability**

The following supplemental data have been deposited at:

<https://datadryad.org/stash/share/2g3gZefPksJaPGILpc8d7gMWqmniILAnQkCI9QRo79c>

Supplemental Data 1: Supp.1.Additional\_Microscopy\_Photos.tar.gz

Supplemental Data 2a: Supp.2a.Physco\_photos.tar.gz

Supplemental Data 2b: Supp.2b.Photo\_Pixel\_Quantification.csv

Supplemental Data 3a: Supp.3a.Physco\_quants.tar.gz

Supplemental Data 3b: Supp.3b.Ppatens\_Berioniawt\_DESeq.csv

Supplemental Data 3c: Supp.3c.Ppatens\_Berioniacu\_DESeq.csv

Supplemental Data 3d: Supp.3d.Ppatens\_Lelongatawt\_DESeq.csv

Supplemental Data 3e: Supp.3e.Ppatens\_Lelongatacu\_DESeq.csv

Supplemental Data 4a: Supp.4a.Perroud\_Mathieu.tar.gz

Supplemental Data 4b: Supp.4b.Perroud\_Mathieu.csv

Supplemental Data 4c: Supp.4c.Significant\_Perroud.csv

Supplemental Data 4d: Supp.4d.Novel\_Expression\_Perroud\_v\_Mathieu.csv

Supplemental Data 4e: Supp.4e.Novel\_Silencing\_Perroud\_v\_Mathieu.csv  
Supplemental Data 5: Supp.5.Media\_Effects\_On\_Ppatens\_PCA.pdf  
Supplemental Data 6: Supp.6.Mortierella\_elongata\_NVP64\_genome\_stats.xlsx  
Supplemental Data 7: Supp.7.SYM\_gene\_hits.txt  
Supplemental Data 8a: Supp.8a.Arabidopsis\_quants.tar.gz  
Supplemental Data 8b: Supp.8b.Algae\_quants.tar.gz  
Supplemental Data 8c: Supp.8c.Orthogroups.tsv  
Supplemental Data 8d: Supp.8d.SingleCopy\_Orthogroups\_all.tsv  
Supplemental Data 8e: Supp.8e.SingleCopy\_Orthogroups\_AT\_PP.tsv  
Supplemental Data 8f: Supp.8f.SingleCopy\_Orthogroups\_CR\_PP.tsv  
Supplemental Data 8g: Supp.8g.Orthogroup\_Significant\_hits\_overlap.xlsx  
Supplemental Data 9: Supp.9.Gene\_Ontology\_Reports.xlsx

The following supplemental code has been deposited at  
<https://zenodo.org/record/8067745>

Supplemental Code 1: Supp.1.Green\_Pixel\_Quantification.ipynb  
Supplemental Code 2a: Supp.2a.Salmon\_analysis.sh  
Supplemental Code 2b: Supp.2b.Split\_files.pl  
Supplemental Code 3: Supp.3.DESeq2\_physco\_v\_Mort.R  
Supplemental Code 4: Supp.4.Perroud\_Mathieu\_comparisons.r  
Supplemental Code 5: Supp.5.Media\_based\_analysis.tar.gz

The genome of *Linnemannia elongata* was submitted to NCBI GenBank and will be accessible as soon as the data is processed, accession number JAXBDG000000000.

### **Conflict of Interest**

The authors declare no conflict of interest.

### **Acknowledgements**

We would like to thank Alan Yocca and Patrick Edger for their recommendations and mentoring in phylogenetics; Aparajita Banerjee, Malik Sankofa, and Balindile Motsa for their assistance in maintaining and growing *P. patens* strains; Sean Johnson, Reid Longley and Natalie VandePol for their help and recommendations for genetic analysis. The work (proposal: 10.46936/10.25585/60001086) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-

05CH11231. DM, GB and BH acknowledge support for this project from the NSF Dimensions of Biodiversity Grant (DEB 1737898) and the NSF-funded doctoral student training grant Integrated training Model in Plant And Compu-Tational Sciences (IMPACTS). BH and GB acknowledges funding from the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409 and support from the Department of Biochemistry and Molecular Biology startup funding and support from AgBioResearch (BH=MICL02454; GB=MICL02416). GB also acknowledges support from the US Department of Energy (DOE) Biological and Environmental Research (BER) and Biological System Science Division (BSSD) under the grant number LANLF59T. BH gratefully acknowledges support from the MSU James K. Billman, Jr. MD endowment. We collectively acknowledge that Michigan State University occupies the ancestral, traditional, and contemporary Lands of the Anishinaabeg – Three Fires Confederacy of Ojibwe, Odawa, and Potawatomi peoples. In particular, the University resides on Land ceded in the 1819 Treaty of Saginaw. We recognize, support, and advocate for the sovereignty of Michigan’s twelve federally-recognized Indian nations, for historic Indigenous communities in Michigan, for Indigenous individuals and communities who live here now, and for those who were forcibly removed from their Homelands. By offering this Land Acknowledgement, we affirm Indigenous sovereignty and will work to hold Michigan State University more accountable to the needs of American Indian and Indigenous peoples.

## REFERENCES

- Akiyama, K., Matsuzaki, K., & Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, 435(7043), Article 7043. <https://doi.org/10.1038/nature03608>
- Alabid, I., Glaeser, S. P., & Kogel, K.-H. (2019). Endofungal Bacteria Increase Fitness of their Host Fungi and Impact their Association with Crop Plants. *Current Issues in Molecular Biology*, 30, 59–74. <https://doi.org/10.21775/cimb.030.059>
- Allan, C., Morris, R. J., & Meisrimler, C.-N. (2022). Encoding, transmission, decoding, and specificity of calcium signals in plants. *Journal of Experimental Botany*, 73(11), 3372–3385. <https://doi.org/10.1093/jxb/erac105>
- Andrews, S. (2010). FastQC A Quality Control tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Ashton, N. W., & Cove, D. J. (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Molecular and General Genetics MGG*, 154(1), 87–95. <https://doi.org/10.1007/BF00265581>
- Bakshi, M., & Oelmüller, R. (2014). WRKY transcription factors. *Plant Signaling & Behavior*, 9, e27700. <https://doi.org/10.4161/psb.27700>
- Berry, J. C., Fahlgren, N., Pokorný, A. A., Bart, R. S., & Veley, K. M. (2018). An automated, high-throughput method for standardizing image color profiles to improve image-based plant phenotyping. *PeerJ*, 6, e5727. <https://doi.org/10.7717/peerj.5727>
- Besserer, A., Puech-Pagès, V., Kiefer, P., Gomez-Roldan, V., Jauneau, A., Roy, S., Portais, J.-C., Roux, C., Bécard, G., & Séjalon-Delmas, N. (2006). Strigolactones Stimulate Arbuscular Mycorrhizal Fungi by Activating Mitochondria. *PLOS Biology*, 4(7), e226. <https://doi.org/10.1371/journal.pbio.0040226>
- Binder, B. M. (2020). Ethylene signaling in plants. *The Journal of Biological Chemistry*, 295(22), 7710–7725. <https://doi.org/10.1074/jbc.REV120.010854>
- Bonfante, P., & Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature Communications*, 1(1), 48. <https://doi.org/10.1038/ncomms1046>
- Bressendorff, S., Azevedo, R., Kenchappa, C. S., Ponce de León, I., Olsen, J. V., Rasmussen, M. W., Erbs, G., Newman, M.-A., Petersen, M., & Mundy, J. (2016). An Innate Immunity Pathway in the Moss *Physcomitrella patens*. *The Plant Cell*, 28(6), 1328–1342. <https://doi.org/10.1105/tpc.15.00774>

- Brzoska, P., & Boos, W. (1988). Characteristics of a *ugp*-encoded and *phoB*-dependent glycerophosphoryl diester phosphodiesterase which is physically dependent on the *ugp* transport system of *Escherichia coli*. *Journal of Bacteriology*, *170*(9), 4125–4135.
- Carapia-Minero, N., Castelán-Vega, J. A., Pérez, N. O., & Rodríguez-Tovar, A. V. (2017). The phosphorelay signal transduction system in *Candida glabrata*: An in-silico analysis. *Journal of Molecular Modeling*, *24*(1), 13. <https://doi.org/10.1007/s00894-017-3545-z>
- Castel, S. E., & Martienssen, R. A. (2013). RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews. Genetics*, *14*(2), 100–112. <https://doi.org/10.1038/nrg3355>
- Causier, B., McKay, M., Hopes, T., Lloyd, J., Wang, D., Harrison, C. J., & Davies, B. (2023). The TOPLESS corepressor regulates developmental switches in the bryophyte *Physcomitrium patens* that were critical for plant terrestrialisation. *The Plant Journal*, *115*(5), 1331–1344. <https://doi.org/10.1111/tpj.16322>
- Chang, Y., Wang, Y., Mondo, S., Ahrendt, S., Andreopoulos, W., Barry, K., Beard, J., Benny, G. L., Blankenship, S., Bonito, G., Cuomo, C., Desiro, A., Gervers, K. A., Hundley, H., Kuo, A., LaButti, K., Lang, B. F., Lipzen, A., O'Donnell, K., ... Spatafora, J. W. (2022). Evolution of zygomycete secretomes and the origins of terrestrial fungal ecologies. *iScience*, *25*(8). <https://doi.org/10.1016/j.isci.2022.104840>
- Chen, J., Gutjahr, C., Bleckmann, A., & Dresselhaus, T. (2015A). Calcium Signaling during Reproduction and Biotrophic Fungal Interactions in Plants. *Molecular Plant*, *8*(4), 595–611. <https://doi.org/10.1016/j.molp.2015.01.023>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen, Y.-E., Cui, J.-M., Su, Y.-Q., Yuan, S., Yuan, M., & Zhang, H.-Y. (2015B). Influence of stripe rust infection on the photosynthetic characteristics and antioxidant system of susceptible and resistant wheat cultivars at the adult plant stage. *Frontiers in Plant Science*, *6*, 779. <https://doi.org/10.3389/fpls.2015.00779>
- Clay, K. (1988). Fungal Endophytes of Grasses: A Defensive Mutualism between Plants and Fungi. *Ecology*, *69*(1), 10–16. <https://doi.org/10.2307/1943155>
- Cove, D. J., Perroud, P.-F., Charron, A. J., McDaniel, S. F., Khandelwal, A., & Quatrano, R. S. (2009). The Moss *Physcomitrella patens*: A Novel Model System for Plant Development and Genomic Studies. *Cold Spring Harbor Protocols*, *2009*(2), pdb.emo115. <https://doi.org/10.1101/pdb.emo115>
- Davey, M. L., Tsuneda, A., & Currah, R. S. (2009). Pathogenesis of bryophyte hosts by the ascomycete *Atracidymella muscivora*. *American Journal of Botany*, *96*(7), 1274–1280. <https://doi.org/10.3732/ajb.0800239>

- de Vries, J., & Archibald, J. M. (2018). Plant evolution: Landmarks on the path to terrestrial life. *New Phytologist*, 217(4), 1428–1434. <https://doi.org/10.1111/nph.14975>
- Delaux, P.-M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., ... Ané, J.-M. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences*, 112(43), 13390–13395. <https://doi.org/10.1073/pnas.1515426112>
- Delaux, P.-M., & Schornack, S. (2021). Plant evolution driven by interactions with symbiotic and pathogenic microbes. *Science*, 371(6531), eaba6605. <https://doi.org/10.1126/science.aba6605>
- Denancé, N., Szurek, B., & Noël, L. D. (2014). Emerging Functions of Nodulin-Like Proteins in Non-Nodulating Plant Species. *Plant and Cell Physiology*, 55(3), 469–474. <https://doi.org/10.1093/pcp/pct198>
- Desirò, A., Hao, Z., Liber, J. A., Benucci, G. M. N., Lowry, D., Roberson, R., & Bonito, G. (2018). Mycoplasma-related endobacteria within Mortierellomycotina fungi: Diversity, distribution and functional insights into their lifestyle. *The ISME Journal*, 12(7), 1743–1757. <https://doi.org/10.1038/s41396-018-0053-9>
- Dickie, I. A., Martínez-García, L. B., Koele, N., Grelet, G.-A., Tylanakis, J. M., Peltzer, D. A., & Richardson, S. J. (2013). Mycorrhizas and mycorrhizal fungal communities throughout ecosystem development. *Plant and Soil*, 367(1), 11–39. <https://doi.org/10.1007/s11104-013-1609-0>
- Douglas, C. M. (2001). Fungal beta(1,3)-D-glucan synthesis. *Medical Mycology*, 39 Suppl 1, 55–66. <https://doi.org/10.1080/mmy.39.1.55.66>
- Du, X., Yang, Z., Ariza, A. J. F., Wang, Q., Xie, G., Li, S., & Du, J. (2022). Structure of plant RNA-DEPENDENT RNA POLYMERASE 2, an enzyme involved in small interfering RNA production. *The Plant Cell*, 34(6), 2140–2149. <https://doi.org/10.1093/plcell/koac067>
- Du, Z.-Y., Zienkiewicz, K., Vande Pol, N., Ostrom, N. E., Benning, C., & Bonito, G. M. (2019). Algal-fungal symbiosis leads to photosynthetic mycelium. *ELife*, 8, e47815. <https://doi.org/10.7554/eLife.47815>
- Duckett, J. G., Carafa, A., & Ligrone, R. (2006). A highly differentiated glomeromycotean association with the mucilage-secreting, primitive antipodean liverwort *Treubia* (Treubiaceae): Clues to the origins of mycorrhizas. *American Journal of Botany*, 93(6), 797–813. <https://doi.org/10.3732/ajb.93.6.797>

- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Eastburn, D. M., McElrone, A. J., & Bilgin, D. D. (2011). Influence of atmospheric and climatic change on plant–pathogen interactions. *Plant Pathology*, 60(1), 54–69. <https://doi.org/10.1111/j.1365-3059.2010.02402.x>
- Feijen, F. A. A., Vos, R. A., Nuytinck, J., & Merckx, V. S. F. T. (2018). Evolutionary dynamics of mycorrhizal symbiosis in land plant diversification. *Scientific Reports*, 8(1), 10698. <https://doi.org/10.1038/s41598-018-28920-x>
- Fonseca, H. M. A. C., & Berbara, R. L. L. (2008). Does *Lunularia cruciata* form symbiotic relationships with either *Glomus proliferum* or *G. intraradices*? *Mycological Research*, 112(9), 1063–1068. <https://doi.org/10.1016/j.mycres.2008.03.008>
- Fontana, D. C., de Paula, S., Torres, A. G., de Souza, V. H. M., Pascholati, S. F., Schmidt, D., & Dourado Neto, D. (2021). Endophytic Fungi: Biological Control and Induced Resistance to Phytopathogens and Abiotic Stresses. *Pathogens*, 10(5), Article 5. <https://doi.org/10.3390/pathogens10050570>
- Friesen, J. A., & Rodwell, V. W. (2004). The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. *Genome Biology*, 5(11), 248. <https://doi.org/10.1186/gb-2004-5-11-248>
- Fürst-Jansen, J. M. R., de Vries, S., & de Vries, J. (2020). Evo-physio: On stress responses and the earliest land plants. *Journal of Experimental Botany*, 71(11), 3254–3269. <https://doi.org/10.1093/jxb/eraa007>
- Gang, H., Li, R., Zhao, Y., Liu, G., Chen, S., & Jiang, J. (2019). Loss of GLK1 transcription factor function reveals new insights in chlorophyll biosynthesis and chloroplast development. *Journal of Experimental Botany*, 70(12), 3125–3138. <https://doi.org/10.1093/jxb/erz128>
- Garcias-Morales, D., Palomar, V. M., Charlot, F., Nogué, F., Covarrubias, A. A., & Reyes, J. L. (2023). N6-Methyladenosine modification of mRNA contributes to the transition from 2D to 3D growth in the moss *Physcomitrium patens*. *The Plant Journal*, 114(1), 7–22. <https://doi.org/10.1111/tpj.16149>
- Giauque, H., & Hawkes, C. V. (2013). Climate affects symbiotic fungal endophyte diversity and performance. *American Journal of Botany*, 100(7), 1435–1444. <https://doi.org/10.3732/ajb.1200568>
- Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., Doust, A. N., Feldman, M. J., Gilbert, K. B., Hodge, J. G., Hoyer, J. S., Lin, A., Liu, S., Lizárraga, C., Lorence, A., Miller, M., Platon, E., Tessman, M., & Sax, T. (2017). PlantCV v2: Image

- analysis software for high-throughput plant phenotyping. *PeerJ*, 5, e4088.  
<https://doi.org/10.7717/peerj.4088>
- Goralogia, G. S., Liu, T.-K., Zhao, L., Panipinto, P. M., Groover, E. D., Bains, Y. S., & Imaizumi, T. (2017). CYCLING DOF FACTOR 1 represses transcription through the TOPLESS co-repressor to control photoperiodic flowering in *Arabidopsis*. *The Plant Journal: For Cell and Molecular Biology*, 92(2), 244–262. <https://doi.org/10.1111/tpj.13649>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), Article 7. <https://doi.org/10.1038/nbt.1883>
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otilar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., & Shabalov, I. (2014). MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Research*, 42(D1), D699–D704. <https://doi.org/10.1093/nar/gkt1183>
- Groenewald, E. G., & Westhuizen, A. J. van der. (1997). Prostaglandins and Related Substances in Plants. *Botanical Review*, 63(3), 199–220.
- Guo, Y., & Narisawa, K. (2018). Fungus-Bacterium Symbionts Promote Plant Health and Performance. *Microbes and Environments*, 33(3), 239–241.  
<https://doi.org/10.1264/jsme2.ME3303rh>
- Hachez, C., Laloux, T., Reinhardt, H., Cavez, D., Degand, H., Grefen, C., De Rycke, R., Inzé, D., Blatt, M. R., Russinova, E., & Chaumont, F. (2014). *Arabidopsis* SNAREs SYP61 and SYP121 coordinate the trafficking of plasma membrane aquaporin PIP2;7 to modulate the cell membrane water permeability. *The Plant Cell*, 26(7), 3132–3147.  
<https://doi.org/10.1105/tpc.114.127159>
- Hall, M. A., Kapuya, J. A., Sivakumaran, S., & John, A. (1977). The role of ethylene in the response of plants to stress. *Pesticide Science*, 8(3), 217–223.  
<https://doi.org/10.1002/ps.2780080307>
- Hanke, S., & Rensing, S. (2010). In vitro association of non-seed plant gametophytes with arbuscular mycorrhiza fungi. *Endocytobiosis and Cell Research*, 20, 95–101.
- Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A., Nakashima, T., Nakajima, M., Mander, L. N., Yamane, H., Hasebe, M., & Nozaki, H. (2010). Endogenous diterpenes derived from *ent*-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss *Physcomitrella patens*. *Plant Physiology*, 153(3), 1085–1097.  
<https://doi.org/10.1104/pp.110.157909>

- Hobbie, E. A., & Boyce, C. K. (2010). Carbon sources for the Palaeozoic giant fungus Prototaxites inferred from modern analogues. *Proceedings of the Royal Society B: Biological Sciences*, 277(1691), 2149–2156. <https://doi.org/10.1098/rspb.2010.0201>
- Hou, Q., Ufer, G., & Bartels, D. (2016). Lipid signalling in plant responses to abiotic stress. *Plant, Cell & Environment*, 39(5), 1029–1048. <https://doi.org/10.1111/pce.12666>
- Hu, Y., Zhong, S., Zhang, M., Liang, Y., Gong, G., Chang, X., Tan, F., Yang, H., Qiu, X., Luo, L., & Luo, P. (2020). Potential Role of Photosynthesis in the Regulation of Reactive Oxygen Species and Defence Responses to *Blumeria graminis* f. Sp. *Tritici* in Wheat. *International Journal of Molecular Sciences*, 21(16), Article 16. <https://doi.org/10.3390/ijms21165767>
- Hunter, L. J. R., Westwood, J. H., Heath, G., Macaulay, K., Smith, A. G., MacFarlane, S. A., Palukaitis, P., & Carr, J. P. (2013). Regulation of RNA-Dependent RNA Polymerase 1 and Isochorismate Synthase Gene Expression in *Arabidopsis*. *PLOS ONE*, 8(6), e66530. <https://doi.org/10.1371/journal.pone.0066530>
- Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K., & Havaux, M. (2003). Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proceedings of the National Academy of Sciences*, 100(8), 4921–4926. <https://doi.org/10.1073/pnas.0736939100>
- Ishida, T., Sugiyama, T., Tabei, N., & Yanagisawa, S. (2014). Diurnal expression of CONSTANS-like genes is independent of the function of cycling DOF factor (CDF)-like transcriptional repressors in *Physcomitrella patens*. *Plant Biotechnology*, 31(4), 293–299. <https://doi.org/10.5511/plantbiotechnology.14.0821a>
- Ivarsson, M., Drake, H., Bengtson, S., & Rasmussen, B. (2020). A Cryptic Alternative for the Evolution of Hyphae. *BioEssays*, 42(6), 1900183. <https://doi.org/10.1002/bies.201900183>
- Ivashuta, S., Liu, J., Liu, J., Lohar, D. P., Haridas, S., Bucciarelli, B., VandenBosch, K. A., Vance, C. P., Harrison, M. J., & Gantt, J. S. (2005). RNA Interference Identifies a Calcium-Dependent Protein Kinase Involved in *Medicago truncatula* Root Development. *The Plant Cell*, 17(11), 2911–2921. <https://doi.org/10.1105/tpc.105.035394>
- Jermey, A. (2011). Soil fungi helped ancient plants to make land. *Nature Reviews Microbiology*, 9(1), Article 1. <https://doi.org/10.1038/nrmicro2494>
- Johnson, J. M., Ludwig, A., Furch, A. C. U., Mithöfer, A., Scholz, S., Reichelt, M., & Oelmüller, R. (2019). The Beneficial Root-Colonizing Fungus *Mortierella hyalina* Promotes the Aerial Growth of *Arabidopsis* and Activates Calcium-Dependent Responses That Restrict *Alternaria brassicae*-Induced Disease Development in Roots. *Molecular Plant-Microbe Interactions®*, 32(3), 351–363. <https://doi.org/10.1094/MPMI-05-18-0115-R>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter

- development team. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows (F. Loizides & B. Schmidt, Eds.; pp. 87–90). IOS Press. <https://doi.org/10.3233/978-1-61499-649-1-87>
- Knack, J. J., Wilcox, L. W., Delaux, P.-M., Ané, J.-M., Piotrowski, M. J., Cook, M. E., Graham, J. M., & Graham, L. E. (2015). Microbiomes of Streptophyte Algae and Bryophytes Suggest That a Functional Suite of Microbiota Fostered Plant Colonization of Land. *International Journal of Plant Sciences*, 176(5), 405–420. <https://doi.org/10.1086/681161>
- Kohler, A., Kuo, A., Nagy, L. G., Morin, E., Barry, K. W., Buscot, F., Canbäck, B., Choi, C., Cichocki, N., Clum, A., Colpaert, J., Copeland, A., Costa, M. D., Doré, J., Floudas, D., Gay, G., Girlanda, M., Henrissat, B., Herrmann, S., ... Martin, F. (2015). Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics*, 47(4), 410–415. <https://doi.org/10.1038/ng.3223>
- Kolosova, N., Miller, B., Ralph, S., Ellis, B. E., Douglas, C., Ritland, K., & Bohlmann, J. (2004). Isolation of high-quality RNA from gymnosperm and angiosperm trees. *BioTechniques*, 36(5), 821–824. <https://doi.org/10.2144/04365ST06>
- Lam, K.-K., LaButti, K., Khalak, A., & Tse, D. (2015). FinisherSC: A repeat-aware tool for upgrading de novo assembly using long reads. *Bioinformatics*, 31(19), 3207–3209. <https://doi.org/10.1093/bioinformatics/btv280>
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R., Vives, C., Morata, J., Symeonidi, A., Hiss, M., Muchero, W., Kamisugi, Y., Saleh, O., Blanc, G., Decker, E. L., ... Rensing, S. A. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, 93(3), 515–533. <https://doi.org/10.1111/tpj.13801>
- Larson, T. J., Ehrmann, M., & Boos, W. (1983). Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the glp regulon. *Journal of Biological Chemistry*, 258(9), 5428–5432. [https://doi.org/10.1016/S0021-9258\(20\)81908-5](https://doi.org/10.1016/S0021-9258(20)81908-5)
- Lehtonen, M. T., Akita, M., Kalkkinen, N., Ahola-Iivarinen, E., Rönholm, G., Somervuo, P., Thelander, M., & Valkonen, J. P. T. (2009). Quickly-released peroxidase of moss in defense against fungal invaders. *New Phytologist*, 183(2), 432–443. <https://doi.org/10.1111/j.1469-8137.2009.02864.x>
- Lehtonen, M. T., Marttinen, E. M., Akita, M., & Valkonen, J. P. T. (2012). Fungi infecting cultivated moss can also cause diseases in crop plants. *Annals of Applied Biology*, 160(3), 298–307. <https://doi.org/10.1111/j.1744-7348.2012.00543.x>
- Li, F., Cheng, C., Cui, F., de Oliveira, M. V. V., Yu, X., Meng, X., Intorne, A. C., Babilonia, K., Li, M., Li, B., Chen, S., Ma, X., Xiao, S., Zheng, Y., Fei, Z., Metz, R. P., Johnson, C. D., Koiwa, H., Sun, W., ... He, P. (2014). Modulation of RNA polymerase II phosphorylation

- downstream of pathogen perception orchestrates plant immunity. *Cell Host & Microbe*, 16(6), 748–758. <https://doi.org/10.1016/j.chom.2014.10.018>
- Liao, H.-L. (2021). The Plant-Growth-Promoting Fungus, *Mortierella elongata*: Its Biology, Ecological Distribution, and Growth-Promoting Activities: SS679/SL466, 3/2021. *EDIS*, 2021(2), Article 2. <https://doi.org/10.32473/edis-ss679-2021>
- Licausi, F., Ohme-Takagi, M., & Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytologist*, 199(3), 639–649. <https://doi.org/10.1111/nph.12291>
- Liepiņa, L. (2012). Occurrence of fungal structures in bryophytes of the boreo-nemoral zone. *Environmental and Experimental Biology*. 10, 35-40.
- Ligrone, R., Carafa, A., Lumini, E., Bianciotto, V., Bonfante, P., & Duckett, J. G. (2007). Glomeromycotean associations in liverworts: A molecular, cellular, and taxonomic analysis. *American Journal of Botany*, 94(11), 1756–1777. <https://doi.org/10.3732/ajb.94.11.1756>
- Loron, C. C., François, C., Rainbird, R. H., Turner, E. C., Borensztajn, S., & Javaux, E. J. (2019). Early fungi from the Proterozoic era in Arctic Canada. *Nature*, 570(7760), Article 7760. <https://doi.org/10.1038/s41586-019-1217-0>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, Y., & Yao, J. (2018). Chloroplasts at the Crossroad of Photosynthesis, Pathogen Infection and Plant Defense. *International Journal of Molecular Sciences*, 19(12), Article 12. <https://doi.org/10.3390/ijms19123900>
- Lutzoni, F., Pagel, M., & Reeb, V. (2001). Major fungal lineages are derived from lichen symbiotic ancestors. *Nature*, 411(6840), 937–940. <https://doi.org/10.1038/35082053>
- Lutzoni, F., Nowak, M. D., Alfaro, M. E., Reeb, V., Miadlikowska, J., Krug, M., Arnold, A. E., Lewis, L. A., Swofford, D. L., Hibbett, D., Hilu, K., James, T. Y., Quandt, D., & Magallón, S. (2018). Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications*, 9(1), 5451. <https://doi.org/10.1038/s41467-018-07849-9>
- Machado, L., Castro, A., Hamberg, M., Bannenberg, G., Gaggero, C., Castresana, C., & de León, I. P. (2015). The *Physcomitrella patens* unique alpha-dioxygenase participates in both developmental processes and defense responses. *BMC Plant Biology*, 15(1), 45. <https://doi.org/10.1186/s12870-015-0439-z>
- Mariconti, L., Pellegrini, B., Cantoni, R., Stevens, R., Bergounioux, C., Cella, R., & Albani, D. (2002). The E2F Family of Transcription Factors from *Arabidopsis thaliana*: NOVEL AND CONSERVED COMPONENTS OF THE RETINOBLASTOMA/E2F PATHWAY IN

PLANTS \*. *Journal of Biological Chemistry*, 277(12), 9911–9919.  
<https://doi.org/10.1074/jbc.M110616200>

- Martin, F., & Nehls, U. (2009). Harnessing ectomycorrhizal genomics for ecological insights. *Current Opinion in Plant Biology*, 12(4), 508–515. <https://doi.org/10.1016/j.pbi.2009.05.007>
- Maughan, S. C., Pasternak, M., Cairns, N., Kiddle, G., Brach, T., Jarvis, R., Haas, F., Nieuwland, J., Lim, B., Müller, C., Salcedo-Sora, E., Kruse, C., Orsel, M., Hell, R., Miller, A. J., Bray, P., Foyer, C. H., Murray, J. A. H., Meyer, A. J., & Cobbett, C. S. (2010). Plant homologs of the *Plasmodium falciparum* chloroquine-resistance transporter, PfCRT, are required for glutathione homeostasis and stress responses. *Proceedings of the National Academy of Sciences*, 107(5), 2331–2336. <https://doi.org/10.1073/pnas.0913689107>
- McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10), 4288–4297. <https://doi.org/10.1093/nar/gks042>
- Mittag, J., Šola, I., Rusak, G., & Ludwig-Müller, J. (2015). *Physcomitrella patens* auxin conjugate synthetase (GH3) double knockout mutants are more resistant to *Pythium* infection than wild type. *Journal of Plant Physiology*, 183, 75–83. <https://doi.org/10.1016/j.jplph.2015.05.015>
- Miyazaki, S., Hara, M., Ito, S., Tanaka, K., Asami, T., Hayashi, K., Kawaide, H., & Nakajima, M. (2018). An Ancestral Gibberellin in a Moss *Physcomitrella patens*. *Molecular Plant*, 11(8), 1097–1100. <https://doi.org/10.1016/j.molp.2018.03.010>
- Miyazaki, S., Katsumata, T., Natsume, M., & Kawaide, H. (2011). The CYP701B1 of *Physcomitrella patens* is an *ent*-kaurene oxidase that resists inhibition by uniconazole-P. *FEBS Letters*, 585(12), 1879–1883. <https://doi.org/10.1016/j.febslet.2011.04.057>
- Modi, A., Vai, S., Caramelli, D., & Lari, M. (2021). The Illumina Sequencing Protocol and the NovaSeq 6000 System. In A. Mengoni, G. Bacci, & M. Fondi (Eds.), *Bacterial Pangenomics: Methods and Protocols* (pp. 15–42). Springer US. [https://doi.org/10.1007/978-1-0716-1099-2\\_2](https://doi.org/10.1007/978-1-0716-1099-2_2)
- Morikawa, T., Saga, H., Hashizume, H., & Ohta, D. (2009). CYP710A genes encoding sterol C22-desaturase in *Physcomitrella patens* as molecular evidence for the evolutionary conservation of a sterol biosynthetic pathway in plants. *Planta*, 229(6), 1311–1322. <https://doi.org/10.1007/s00425-009-0916-4>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C. H., Yang, Z., Schneider, H., & Donoghue, P. C. J. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences*, 115(10), E2274–E2283. <https://doi.org/10.1073/pnas.1719588115>

- Mueller, M. J. (1998). Radically novel prostaglandins in animals and plants: The isoprostanes. *Chemistry & Biology*, 5(12), R323–R333. [https://doi.org/10.1016/S1074-5521\(98\)90660-3](https://doi.org/10.1016/S1074-5521(98)90660-3)
- Müller, H., Bracken, A. P., Vernell, R., Moroni, M. C., Christians, F., Grassilli, E., Prosperini, E., Vigo, E., Oliner, J. D., & Helin, K. (2001). E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes & Development*, 15(3), 267–285. <https://doi.org/10.1101/gad.864201>
- Nakano, Y., Yamaguchi, M., Endo, H., Rejab, N. A., & Ohtani, M. (2015). NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. *Frontiers in Plant Science*, 0. <https://doi.org/10.3389/fpls.2015.00288>
- Naumann, M., Schübler, A., & Bonfante, P. (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *The ISME Journal*, 4(7), Article 7. <https://doi.org/10.1038/ismej.2010.21>
- Nelsen, M. P., Lücking, R., Boyce, C. K., Lumbsch, H. T., & Ree, R. H. (2020). No support for the emergence of lichens prior to the evolution of vascular plants. *Geobiology*, 18(1), 3–13. <https://doi.org/10.1111/gbi.12369>
- Nguyen, T. T. T., Park, S. W., Pangging, M., & Lee, H. B. (2019). Molecular and Morphological Confirmation of Three Undescribed Species of *Mortierella* from Korea. *Mycobiology*, 47(1), 31–39. <https://doi.org/10.1080/12298093.2018.1551854>
- Ohshima, S., Sato, Y., Fujimura, R., Takashima, Y., Hamada, M., Nishizawa, T., Narisawa, K., & Ohta, H. 2016. (n.d.). *Mycoavidus cysteinexigens* gen. Nov., sp. Nov., an endohyphal bacterium isolated from a soil isolate of the fungus *Mortierella elongata*. *International Journal of Systematic and Evolutionary Microbiology*, 66(5), 2052–2057. <https://doi.org/10.1099/ijsem.0.000990>
- Otero-Blanca, A., Pérez-Llano, Y., Reboledo-Blanco, G., Lira-Ruan, V., Padilla-Chacon, D., Folch-Mallol, J. L., Sánchez-Carbente, M. D. R., Ponce De León, I., & Batista-García, R. A. (2021). *Physcomitrium patens* Infection by *Colletotrichum gloeosporioides*: Understanding the Fungal–Bryophyte Interaction by Microscopy, Phenomics and RNA Sequencing. *Journal of Fungi*, 7(8), 677. <https://doi.org/10.3390/jof7080677>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon: Fast and bias-aware quantification of transcript expression using dual-phase inference. *Nature Methods*, 14(4), 417–419. <https://doi.org/10.1038/nmeth.4197>
- Perroud, P.-F., Haas, F. B., Hiss, M., Ullrich, K. K., Alboresi, A., Amirebrahimi, M., Barry, K., Bassi, R., Bonhomme, S., Chen, H., Coates, J. C., Fujita, T., Guyon-Debast, A., Lang, D., Lin, J., Lipzen, A., Nogué, F., Oliver, M. J., Ponce de León, I., ... Rensing, S. A. (2018). The *Physcomitrella patens* gene atlas project: Large-scale RNA-seq based expression data. *The Plant Journal*, 95(1), 168–182. <https://doi.org/10.1111/tpj.13940>

- Ponce de León, I. (2011). The Moss *Physcomitrella patens* as a Model System to Study Interactions between Plants and Phytopathogenic Fungi and Oomycetes. *Journal of Pathogens*, 2011, e719873. <https://doi.org/10.4061/2011/719873>
- Ponce De León, I., Schmelz, E. A., Gaggero, C., Castro, A., Álvarez, A., & Montesano, M. (2012). *Physcomitrella patens* activates reinforcement of the cell wall, programmed cell death and accumulation of evolutionary conserved defence signals, such as salicylic acid and 12-oxo-phytodienoic acid, but not jasmonic acid, upon *Botrytis cinerea* infection. *Molecular Plant Pathology*, 13(8), 960–974. <https://doi.org/10.1111/j.1364-3703.2012.00806.x>
- Proust, H., Hoffmann, B., Xie, X., Yoneyama, K., Schaefer, D. G., Yoneyama, K., Nogué, F., & Rameau, C. (2011). Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*. *Development*, 138(8), 1531–1539. <https://doi.org/10.1242/dev.058495>
- Raetz, C. R. H. (1986). Molecular Genetics of Membrane Phospholipid Synthesis. *Annual Review of Genetics*, 20(1), 253–291. <https://doi.org/10.1146/annurev.ge.20.120186.001345>
- Rathgeb, U., Chen, M., Buron, F., Feddermann, N., Schorderet, M., Raisin, A., Häberli, G.-Y., Marc-Martin, S., Keller, J., Delaux, P.-M., Schaefer, D. G., & Reinhardt, D. (2020). VAPYRIN-like is required for development of the moss *Physcomitrella patens*. *Development*, 147(11). <https://doi.org/10.1242/dev.184762>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198. <https://doi.org/10.1093/nar/gkz369>
- Read, D. J., Duckett, J. G., Francis, R., Ligrone, R., & Russell, A. (2000). Symbiotic fungal associations in “lower” land plants. Royal Society. <https://doi.org/10.1098/rstb.2000.0617>
- Reboledo, G., Agorio, A. d., Vignale, L., Batista-García, R. A., & Ponce De León, I. (2021). Transcriptional profiling reveals conserved and species-specific plant defense responses during the interaction of *Physcomitrium patens* with *Botrytis cinerea*. *Plant Molecular Biology*, 107(4), 365–385. <https://doi.org/10.1007/s11103-021-01116-0>
- Reboledo, G., Agorio, A., Vignale, L., Batista-García, R. A., & Ponce De León, I. (2020). *Botrytis cinerea* Transcriptome during the Infection Process of the Bryophyte *Physcomitrium patens* and Angiosperms. *Journal of Fungi*, 7(1), 11. <https://doi.org/10.3390/jof7010011>
- Reimand, J., Kull, M., Peterson, H., Hansen, J., & Vilo, J. (2007). g:Profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Research*, 35(Web Server issue), W193–W200. <https://doi.org/10.1093/nar/gkm226>

- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S.-I., ... Boore, J. L. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science (New York, N.Y.)*, 319(5859), 64–69.  
<https://doi.org/10.1126/science.1150646>
- Reski, R., & Abel, W. O. (1985). Induction of budding on chloronemata and caulonemata of the moss, *Physcomitrella patens*, using isopentenyladenine. *Planta*, 165(3), 354–358.  
<https://doi.org/10.1007/BF00392232>
- Roberts, A., Roberts, E., & Haigler, C. (2012). Moss cell walls: Structure and biosynthesis. *Frontiers in Plant Science*, 3, 166. <https://doi.org/10.3389/fpls.2012.00166>
- Roberts, A. W., Lahnstein, J., Hsieh, Y. S. Y., Xing, X., Yap, K., Chaves, A. M., Scavuzzo-Duggan, T. R., Dimitroff, G., Lonsdale, A., Roberts, E., Bulone, V., Fincher, G. B., Doblin, M. S., Bacic, A., & Burton, R. A. (2018). Functional Characterization of a Glycosyltransferase from the Moss *Physcomitrella patens* Involved in the Biosynthesis of a Novel Cell Wall Arabinoglucan. *The Plant Cell*, 30(6), 1293–1308.  
<https://doi.org/10.1105/tpc.18.00082>
- Russell, J., & Bulman, S. (2005). The liverwort *Marchantia foliacea* forms a specialized symbiosis with arbuscular mycorrhizal fungi in the genus *Glomus*. *New Phytologist*, 165(2), 567–579. <https://doi.org/10.1111/j.1469-8137.2004.01251.x>
- Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P., & Bonfante, P. (2016). Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *The ISME Journal*, 10(1), 130–144.  
<https://doi.org/10.1038/ismej.2015.91>
- Savage, Z., Duggan, C., Toufexi, A., Pandey, P., Liang, Y., Segretin, M. E., Yuen, L. H., Gaboriau, D. C. A., Leary, A. Y., Tumtas, Y., Khandare, V., Ward, A. D., Botchway, S. W., Bateman, B. C., Pan, I., Schattat, M., Sparkes, I., & Bozkurt, T. O. (2021). Chloroplasts alter their morphology and accumulate at the pathogen interface during infection by *Phytophthora infestans*. *The Plant Journal*, 107(6), 1771–1787.  
<https://doi.org/10.1111/tpj.15416>
- Scharte, J., Schön, H., & Weis, E. (2005). Photosynthesis and carbohydrate metabolism in tobacco leaves during an incompatible interaction with *Phytophthora nicotianae*. *Plant, Cell & Environment*, 28(11), 1421–1435. <https://doi.org/10.1111/j.1365-3040.2005.01380.x>
- Simkin, A. J., Guirimand, G., Papon, N., Courdavault, V., Thabet, I., Ginis, O., Bouzid, S., Giglioli-Guivarc'h, N., & Clastre, M. (2011). Peroxisomal localisation of the final steps of the mevalonic acid pathway in planta. *Planta*, 234(5), 903–914.  
<https://doi.org/10.1007/s00425-011-1444-6>

- Smith, S. E., & Read, D. J. (2010). *Mycorrhizal Symbiosis*. Academic Press.
- Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences (4:1521). *F1000Research*. <https://doi.org/10.12688/f1000research.7563.2>
- Soufi, A., & Jayaraman, P.-S. (2008). PRH/Hex: An oligomeric transcription factor and multifunctional regulator of cell fate. *The Biochemical Journal*, *412*(3), 399–413. <https://doi.org/10.1042/BJ20080035>
- Stenzel, I., Hause, B., Miersch, O., Kurz, T., Maucher, H., Weichert, H., Ziegler, J., Feussner, I., & Wasternack, C. (2003). Jasmonate biosynthesis and the allene oxide cyclase family of *Arabidopsis thaliana*. *Plant Molecular Biology*, *51*(6), 895–911. <https://doi.org/10.1023/A:1023049319723>
- Su, C. (2023). Pectin modifications at the symbiotic interface. *New Phytologist*, *238*(1), 25–32. <https://doi.org/10.1111/nph.18705>
- Swarbrick, P. J., Schulze-Lefert, P., & Scholes, J. D. (2006). Metabolic consequences of susceptibility and resistance (race-specific and broad-spectrum) in barley leaves challenged with powdery mildew. *Plant, Cell & Environment*, *29*(6), 1061–1076. <https://doi.org/10.1111/j.1365-3040.2005.01472.x>
- Takahashi, T., & Kakehi, J.-I. (2010). Polyamines: Ubiquitous polycations with unique roles in growth and stress responses. *Annals of Botany*, *105*(1), 1–6. <https://doi.org/10.1093/aob/mcp259>
- Tominaga, M., Kojima, H., Yokota, E., Nakamori, R., Anson, M., Shimmen, T., & Oiwa, K. (2012). Calcium-induced mechanical change in the neck domain alters the activity of plant myosin XI. *The Journal of Biological Chemistry*, *287*(36), 30711–30718. <https://doi.org/10.1074/jbc.M112.346668>
- Uehling, J., Gryganskyi, A., Hameed, K., Tschaplinski, T., Misztal, P. K., Wu, S., Desirò, A., Vande Pol, N., Du, Z., Zienkiewicz, A., Zienkiewicz, K., Morin, E., Tisserant, E., Splivallo, R., Hainaut, M., Henrissat, B., Ohm, R., Kuo, A., Yan, J., ... Bonito, G. (2017). Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycoavidus cysteinexigens*. *Environmental Microbiology*, *19*(8), 2964–2983. <https://doi.org/10.1111/1462-2920.13669>
- Vandepol, N., Liber, J., Yocca, A., Matlock, J., Edger, P., & Bonito, G. (2022). *Linnemannia elongata* (Mortierellaceae) stimulates *Arabidopsis thaliana* aerial growth and responses to auxin, ethylene, and reactive oxygen species. *PLOS ONE*, *17*(4), e0261908. <https://doi.org/10.1371/journal.pone.0261908>

- Vega-Sánchez, M. E., Verherbruggen, Y., Scheller, H. V., & Ronald, P. C. (2013). Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signaling & Behavior*, 8(2), e23143. <https://doi.org/10.4161/psb.23143>
- Verma, A., Shameem, N., Jatav, H. S., Sathyanarayana, E., Parray, J. A., Poczai, P., & Sayyed, R. Z. (2022). Fungal Endophytes to Combat Biotic and Abiotic Stresses for Climate-Smart and Sustainable Agriculture. *Frontiers in Plant Science*, 13. <https://www.frontiersin.org/articles/10.3389/fpls.2022.953836>
- Voiniciuc, C., Pauly, M., & Usadel, B. (2018). Monitoring Polysaccharide Dynamics in the Plant Cell Wall [OPEN]. *Plant Physiology*, 176(4), 2590–2600. <https://doi.org/10.1104/pp.17.01776>
- Waller, K. L., Muhle, R. A., Ursos, L. M., Horrocks, P., Verdier-Pinard, D., Sidhu, A. B. S., Fujioka, H., Roepe, P. D., & Fidock, D. A. (2003). Chloroquine Resistance Modulated *in Vitro* by Expression Levels of the *Plasmodium falciparum* Chloroquine Resistance Transporter \*. *Journal of Biological Chemistry*, 278(35), 33593–33601. <https://doi.org/10.1074/jbc.M302215200>
- Wang, X., & He, Y. (2015). Tissue Culturing and Harvesting of Protonemata from the Moss *Physcomitrella patens*. *BIO-PROTOCOL*, 5(15). <https://doi.org/10.21769/BioProtoc.1556>
- Wang, F., Lai, L., Liu, Y., Yang, B., & Wang, Y. (2016). Expression and Characterization of a Novel Glycerophosphodiester Phosphodiesterase from *Pyrococcus furiosus* DSM 3638 That Possesses Lysophospholipase D Activity. *International Journal of Molecular Sciences*, 17(6), 831. <https://doi.org/10.3390/ijms17060831>
- Wang, J., Lian, N., Zhang, Y., Man, Y., Chen, L., Yang, H., Lin, J., & Jing, Y. (2022). The Cytoskeleton in Plant Immunity: Dynamics, Regulation, and Function. *International Journal of Molecular Sciences*, 23(24), Article 24. <https://doi.org/10.3390/ijms232415553>
- Waters, M. T., Scaffidi, A., Flematti, G. R., & Smith, S. M. (2013). The origins and mechanisms of karrikin signalling. *Current Opinion in Plant Biology*, 16(5), 667–673. <https://doi.org/10.1016/j.pbi.2013.07.005>
- Yasumura, Y., Moylan, E. C., & Langdale, J. A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *The Plant Cell*, 17(7), 1894–1907. <https://doi.org/10.1105/tpc.105.033191>
- Wei, Q., Wang, W., Hu, T., Hu, H., Mao, W., Zhu, Q., & Bao, C. (2018). Genome-wide identification and characterization of Dof transcription factors in eggplant (*Solanum melongena* L.). *PeerJ*, 6, e4481. <https://doi.org/10.7717/peerj.4481>
- White Jr, J. F., & Torres, M. S. (2010). Is plant endophyte-mediated defensive mutualism the result of oxidative stress protection? *Physiologia Plantarum*, 138(4), 440–446. <https://doi.org/10.1111/j.1399-3054.2009.01332.x>

- Xia, W., Yu, H., Cao, P., Luo, J., & Wang, N. (2017). Identification of TIFY Family Genes and Analysis of Their Expression Profiles in Response to Phytohormone Treatments and *Melampsora larici-populina* Infection in Poplar. *Frontiers in Plant Science*, 8. <https://www.frontiersin.org/articles/10.3389/fpls.2017.00493>
- Yang, Z., Duan, L., Li, H., Tang, T., Chen, L., Hu, K., Yang, H., & Liu, L. (2022). Regulation of Heat Stress in *Physcomitrium (Physcomitrella) patens* Provides Novel Insight into the Functions of Plant RNase H1s. *International Journal of Molecular Sciences*, 23(16), Article 16. <https://doi.org/10.3390/ijms23169270>
- Yasumura, Y., Moylan, E. C., & Langdale, J. A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *The Plant Cell*, 17(7), 1894–1907. <https://doi.org/10.1105/tpc.105.033191>
- Yin, Y., Vafeados, D., Tao, Y., Yoshida, S., Asami, T., & Chory, J. (2005). A New Class of Transcription Factors Mediates Brassinosteroid-Regulated Gene Expression in *Arabidopsis*. *Cell*, 120(2), 249–259. <https://doi.org/10.1016/j.cell.2004.11.044>
- Zhao, S., Ye, Z., & Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*, rna.074922.120. <https://doi.org/10.1261/rna.074922.120>
- Zhang, K., Bonito, G., Hsu, C.-M., Hameed, K., Vilgalys, R., & Liao, H.-L. (2020). *Mortierella elongata* Increases Plant Biomass among Non-Leguminous Crop Species. *Agronomy*, 10(5), Article 5. <https://doi.org/10.3390/agronomy10050754>

**CHAPTER 3**  
**Rule-Based Deconstruction and Reconstruction of the Diterpene Library: A Simulation of  
Synthesis and Unravelling of Compound Structural Diversity**

Mathieu, D., Schlecht, S., van Aalst, M., Shebek, K., Busta, L., Babineau, N., Ebenhöf, O.,  
Hamberger, B.

## **Abstract**

Terpenoids make up the largest class of specialized metabolites with over 180,000 reports currently across all kingdoms of life. Their synthesis accentuates one of nature's most choreographed enzymatic and non-reversible chemistries, leading to an extensive range of structural functionality and diversity. Current terpenoid repositories provide a seemingly endless playground of information regarding structure, sourcing, and synthesis. Efforts here investigate entries for the 20-carbon diterpenoid variants and deconstruct the complex patterns into simple, categorical groups. This deconstruction approach reduces over 60,000 unique compound entries down to less than 1,000 categorical structures. Furthermore, over 75% of all diversity can be represented by just 25 total structures. The diversity of diterpenoids could be identified at an atom-by-atom comparison, across the total compound landscape, and distributed throughout the tree of life. Additionally, these core structures provide guidelines for predicting how this diversity first originates via the help of diterpene synthases. Over 95% of all diterpenoid structures rely on cyclization. Here a reconstructive approach is reapplied to the data based on known biochemical rules to model the birth of compound diversity. This computational synthesis validates previously identified reaction products and pathways, as well as predicts multiple trajectories for synthesizing real and theoretical compounds. This deconstructive and reconstructive approach applied to the diterpene landscape provides modular, flexible, and an easy-to-use toolset for categorically simplifying otherwise complex or hidden patterns.

## **Keywords**

Carbocation, Diterpene, Compound Libraries, Computational Modelling, Rule-Based Analysis

## **Significance Statement**

Analyses performed here categorically curate compounds, provide user-friendly software with modularity as part of the design, and views the diterpene landscape from new perspectives. Compounds are investigated based on origin, abundance, synthesis, and diversification. Full data analysis offers a contrasting yet synergistic perspective, enabling the exploration of both structural complexities and categorical simplicities, while still considering the individual compound(s) and how they fit within the complete diterpene landscape.

## **Introduction**

Terpenoids are specialized metabolites known for their metabolic diversity, expansive utility, and vast abundance throughout the tree of life. They are structurally characterized by the combination of the 5-carbon molecules, isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP), to form the linear prenyl-diphosphates, which are further linked to form the mono- (C10), sesqui- (C15), di- (C20), sester- (C25), tri- (C30), tetra- (C40) and polyterpenoid classes. Currently, over 180,000 unique terpenoid structures have been reported in the Dictionary of Natural Products (DNP) and TeroKit databases [Buckingham 2015, Zeng et al. 2020, Buckingham 2023]. Most reported structural diversity originates from the cyclization of linear precursors in a number of different ways via terpene synthase (TPS) activity [Degenhardt et al., 2009; Durairaj et al., 2019; Johnson et al., 2019a; Miller et al. 2020; Zeng et al., 2019] and through further modification via oxidative functionalization by NADPH-dependent cytochrome P450 mono-oxygenases, subsequently acting dehydrogenases, 2-oxoglutarate dependent oxygenases, or a range of transferase affording conjugates [Kawai et al. 2014; Johnson et al.,

2019b; Karunanithi and Zerbe, 2019; Luo et al. 2016; Miller et al. 2020; Pateraki et al., 2017]. In nature, these compounds function with purpose in defense, pollinator attraction, developmental signaling, and interspecies communication [Degenhardt et al. 2003; Theis and Lerdau 2003; Aros et al. 2012; Boncan et al. 2020; Caissard et al. 2004; Chou et al. 2023; Cseke et al. 2007; Gershenzon & Dudareva 2007; Dötterl & Gershenzon 2023; Erbilgin et al., 2006; Jassbi et al. 2008; Heiling et al. 2010; Huang & Osbourn 2019; Keeling and Bohlmann, 2006; Laurent et al. 2003; Li et al. 2021; Lipińska et al. 2022; Lu et al. 2018; Miller et al. 2020; Miyazaki et al. 2015; Nagel et al. 2014; Ndi et al. 2007; Piccoli & Bottini 2013; Proffit et al. 2020; Schiebe et al., 2012; Toyomasu et al. 2014; Wang et al. 2023; Zhao et al., 2011]. The intrinsic role terpenoids serve in communication and defense likely has been a driving force for the observed diversity to date [Luo et al. 2012; Chen et al. 2014; Li et al. 2021; Zhang et al. 2022]. Additionally, because plants and other sessile organisms are unable to move from threats, there is greater pressure for these responses to be chemical [Weng 2013; Villegas-Plazas et al. 2018; Weng et al. 2021]. From a human perspective, terpenes offer broad applications as fuels, pharmaceuticals, nutraceuticals, fragrances, and pesticides [Degenhardt et al. 2003; González-Coloma et al. 2014; Hausch et al. 2015; Koul 2008; Lange et al. 2011; Schalk et al. 2011; Phillipe et al. 2014; Celedon & Bohlmann 2016; Kut yana & Bornemann 2018; Nuutinen 2018; Tetali 2019; Wang et al. 2005; Wani et al. 1971; Wilson & Roberts 2011; Zerbe et al. 2012; Zerbe and Bohlmann 2014; Zhao et al. 2016].

Work presented here utilizes diterpenes as a case study to showcase developed software and reveal complex patterns that exist within the current structural landscape. Over 95% of diterpenes are cyclized either sequentially by a class II diTPS followed by a class I diTPS, or by a class I TPS alone. Class II/class I derived diterpenes most commonly represent structures with

a characteristic decalin-core and are blanket referred to here as “labdane-derived”. These compounds are synthesized by an initial protonation at the GGDP tail (class II), which then leads to cationic cycloisomerization and subsequent quenching by deprotonation or hydroxylation. This is followed by the removal of the diphosphate (class I), forming another carbocation that can also cyclize (as in the case of kaurenes) and/or be quenched (as in the case of labdanes) [Peters 2010; Tantillo 2010]. The other most common synthesis, catalyzed by class I TPS, forms a carbocation by removing the diphosphate, which can lead to larger ring formations. This mode of synthesis is referred to here as “macrocyclic-derived”. Class II/class I and class I chemistries that orchestrate diterpene synthesis provide the foundation for most diterpene structural diversity. Studying the immense size and scale of the current terpene landscape requires analysis to be performed in a computational space. This work has been done previously through machine learning and/or rule-based approaches to mirror biochemical reactions [Tantillo 2010, Tantillo 2011, Durairaj et al. 2021; Hosseini & Pereira 2023; Shebek et al. 2023; Strutz et al. 2022; Zeng et al. 2019; Zeng et al. 2020; Zeng et al. 2022]. Machine learning can generate an abundance of compounds and parse out complex patterns within datasets, but also faces limitations where biases, and errors within a dataset may not be readily detected, which affect interpretation and reproducibility [Carbone 2022; Malik 2020]. Alternatively, a rule-based approach provides a higher degree of control to the user, allows the ability for modularity based on conditions, grants ease to accommodate a growing database in the future, and emphasizes human consideration when modelling the metabolic landscape. Here, rule-based methods are utilized with Simplified Molecular Input Line Entry System (SMILES) as inputs, which are the presentation of chemical structures as text in a computational space [Kumar 2021; Toropov et al. 2005; Weininger 1988; Weininger et al. 1989; Weininger 1990]. The reaction rules applied are represented in a SMILES

Arbitrary Target Specification (SMARTS) format for pattern recognition within compound structures and determining if a reaction is permissible [Arteca 1996; Landrum 2023; O'Donnell et al. 1991; Sayle 1997; Todeschini & Consonni 2003; Van Drie et al. 1989].

The rule-based methodology here operates on the reported diterpenes in the DNP (>25K; v30.1) and TeroKit (>40K; v2.0) databases to uncover complex patterns regarding diterpene structural diversity, synthesis, and origin. Diterpene synthases catalyze complex, multi-step chemistries derived from diphosphate cleavage, protonation, carbocation rearrangements via nucleophilic attacks, methyl and hydride shifts, and eventual quenching and resolution of carbocations. A reverse (deconstruction) and forward (synthesis) approach are used here to model diterpene biochemistry targeting the synthesis of diterpene backbones and skeletons. The derived backbones are defined here as only the portion of the molecule originating from a prenyl-diphosphate but with retained stereochemistry, bond information, and hydroxyl substituted R-groups. Diterpene skeletons are defined as the most simplified 20-carbon structures limited to carbon-carbon linkages, without considering bond variation, stereochemistry, or hydroxyl groups. The reverse approach isolates and identifies these diterpene backbones and skeletons. The forward approach predicts diterpene backbone and skeleton formation using carbocation reactions mirroring the mechanisms of class II and class I diTPSs [Shebek et al. 2023; Strutz et al. 2022]. This reconstructive modelling approach provides a unique platform to demonstrate known carbocation rearrangements and produce an output of stepwise mechanisms for both known and theoretical diterpene chemistries. Deconstructive and reconstructive paired methods provide a unique synergy for simplifying the abundance of reported compounds, examining the origin of diversity, and predicting mechanisms of synthesis.

## Methods & Materials

### *Deconstruction of the DNP and TeroKit diterpene libraries to skeletons and backbones*

The DNP database<sup>8</sup> (v30.2) was queried within the rubric ‘Type of Compound Words’ for “diterpen”. Database hits were semi-automatically downloaded, extracted, and concatenated, collecting information on Chemical name, formula, weight, SMILE, InChi, compound type, and biological source for each compound (Supplemental Data 1a). The TeroKit molecule database<sup>9</sup> (v2.0) was downloaded and subsequently parsed into “sesquiterpene”, “diterpene”, and “triterpene” datasets using the grep command (Supplemental Data 1b-d) [Zeng et al. 2020; Zeng et al. 2022]. These datasets included TeroKit compound ID, molecular formula, InChi, and SMILES. SMILES from each datasets were used as input for deconstruction. Python (v3.9) was used with libraries, pandas (v1.5.1), numpy (v1.25.0), re (v2.2.1), matplotlib (v3.7.1), and RDKit (v2022.09.1) (Supplemental Code 1) [Landrum 2023]. Formatted SMILES from each database were subsequently and iteratively deconstructed to backbones and skeletons. The first function removed any portion of the molecule containing: boron, halogens, silicon, phosphate, sulfur, selenium, tin, fatty acids, saccharides, coumarin, nitrogenous bases, other nitrogen containing R groups, ester linked R groups, ether linked R groups, and other “non-terpene” derived carbon side chains. Instances of isotopic carbon or charged carbon atoms were also converted to a neutral <sup>12</sup>C. The output returned a backbone with any R-group substituted with an alcohol moiety. Bond and stereochemical variation was retained, along with a list of each step taken to deconstruct the original compound. These backbones were then flattened to a carbon skeleton by converting all covalent bonds to single bonds and removing hydroxyl groups, resulting in exclusively carbon-carbon connections. Backbones, skeletons, and final carbon numbers were

---

<sup>8</sup> <http://dnp.chemnetbase.com>

<sup>9</sup> <http://terokit.qmclab.com/data.html>

appended in order to update original datasets (Supplemental Data 2a-2d) and generate abundance summaries based on shared skeleton structures for the DNP (Supplemental Data 2e) and TeroKit (Supplemental Data 2f) datasets. Deconstruction of sesquiterpene and triterpene databases were performed as proof-of-functionality and to support robustness of programs (Supplemental Data 2c, 2d). Representation of TeroKit skeletons were visualized based on abundance to identify percentile coverages for the total diterpenoid database (Supplemental Data 2g). Carbon number of deconstructed diterpenoids identified database outliers, database entry errors, and non-C<sub>20</sub> diterpenoids, such as cleavage products and apo-terpenoids (Supplemental Data 2h, 2i).

### ***Structural comparison of the DNP diterpene skeletons***

Due to a higher degree of curation, skeletons identified from the DNP were used for structural comparison (Supplemental Data 2e). A comparison matrix was generated by converting the 671 skeleton SMILES to bit vectors (binary vectors) using RDKit (v2022.09.1) [Landrum 2023]. All skeleton comparisons were made based on bit vector comparison calculations to determine similarity scores ranging from 0 to 1 and to create a 671x671 matrix (Supplemental Data 3). This matrix was visualized with seaborn (v0.12.2) as a heatmap (Figure 3.1; Supplemental Code 2). A principal component analysis (PCA) was conducted on this matrix using sklearn (v1.1.3) functions for transformation, PCA generation, and variance calculations (Supplemental Code 2).

### ***Modelling diterpene carbocation reactions for backbone generation in Pickaxe***

The software, Pickaxe [Shebek et al. 2023], was modified for job submission in a bash terminal (Supplemental Code 3a). The following Pickaxe settings were changed from their default: “input\_cpds”, “product\_cpds”, “output\_dir”, “generations”, “sample\_size=1000000”, “coreactant\_list”, “rules\_list”, “processes=24”, “kekulize=False”, “processes=24”,

“quiet=False”, “neutralise=False”. Quality of life additions were made to print an abundance summary of rule usage and to include date and time in output naming. Because Pickaxe has multiple checks for validating compound structures, the inclusion of charged carbocation atoms required adaptation of the rules. In brief, the noble gas, Xenon, was used to represent carbocations in all reactions instead. The creation of diTPS SMARTs rules had three guiding principles. First, the majority of carbocation rules prioritized the formation of reactive centers on tertiary carbons, because of the higher commonality and stability of carbocations at these positions [Tantillo 2010]. Generally, these rules represented a nucleophilic attack of a double bond, ring formation, and subsequent generation of a new tertiary carbocation. Second, when possible, reaction rules focused on the ‘reactive unit’ as opposed to the entire molecule. This maintained specificity of known reactions and allowed the same rule to be more generalizable in other conditions as well, reducing redundancy within the ruleset. Third, high specificity was used for instances where rules fulfilled niche reactions, such as in the case of concerted reactions, carbocation stabilization via delocalization or hyperconjugation, or the generation of secondary carbocations [Tantillo 2011]. These respective rules would be generated to resemble the identified mechanism as closely as possible within the context of Pickaxe. Illustrations of specific reaction rules and their sourcing can be found in Supplemental Data 4.

Pickaxe had four separate submissions where 1.) represented unfiltered class II diTPS action, 2.) represented unfiltered class I diTPS action, 3.) represented the class II and class I rules from steps 1 and 2 but now filtered to known diterpene backbones and 4.) explored where remaining skeleton diversity may be originating. Updated coreactant lists and custom biochemical rulesets were generated for each of the four iterations (Supplemental Data 5a-d). The first iteration had 10 generations of reactions and used geranylgeranyl diphosphate (GGDP) as the initial substrate.

This iteration used the class II diTPS reaction ruleset, specifically excluding diphosphate cleavage and any rules generating macrocyclic compounds (Supplemental Data 5a). The second iteration had 10 generations of reactions, used GGDP and all products obtained from the first iteration of compounds. The second iteration used the class I diTPS reaction ruleset (Supplemental Data 5b). All compounds generated from the first and second iteration not containing a diphosphate or carbocation were converted to carbon skeletons, converted to a canonical SMILE format, and compared to all previously identified C20 skeletons (Supplemental Data 2e, 2f). If a compound matched a previously identified skeleton, the original compound was marked as a successful target match, representing 80 backbones (736 redundant structures) in total (Supplemental Code 3c). First and second Pickaxe submissions quantified total novel compounds synthesized at each generational step to track overall abundance overtime (Figure 3.2a).

Validation of the model necessitates comparison to known terpene synthesis reactions with the third iteration, which was pruned based on successful targets generated from class I activity (Figure 3.2b; Supplemental Data 5C, 6). All filtered reactions were concatenated into a single file and converted into a network containing information on edge (reactions) and node (compound) data between compounds (Supplemental Code 3c) along with key identifiers like, SMILES, and compound classifications (GGDP, class II intermediates, class II products, class I intermediates, final targets). These were made into a network in Cytoscape (v3.10.1) to represent known diTPS carbocation reactions (Figure 3.2b; Supplemental Data 6) [Shannon et al. 2003]. Within Cytoscape the “analyze network function” was performed to evaluate stress centrality distribution and betweenness centrality. Reactions specific for generating Kaurene and Taxadiene synthesis were also analyzed independently in Cytoscape (Figure 3.2c, 3.2d; Supplemental Data

6) [Shannon et al. 2003]. The fourth Pickaxe submission performed 3 generations of rules that further modified skeleton structure in ways beyond the scope of diTPS activity. These rules broke carbon rings, created carbon rings, expanded/collapsed connecting rings, and shifted carbon side chains (Supplemental Data 5d).

### ***Predicting carbocation quenching mechanisms for common diterpene backbones***

The top 20 most common diterpenes in the TeroKit database (Supplemental Data 2b, 2f) were used to predict carbocation resolution via rearrangement or quenching with water (Supplemental Code 4). Compound lists also identified aromatic ring systems and predicted post TPS activity on double bond desaturation and saturation post carbocation cascades. Within a skeleton class, all backbones had hydroxyl groups removed and the remaining number of hydrogen atoms were counted. Compounds with 32, 34, or 36 hydrogen atoms present were determined to represent the resolution of zero, one, or two carbocations via rearrangement or quenching with water respectively. Class II/class I derived compounds generate 2 carbocations, whereas class I derived compounds generate 1 carbocation, and the phytanes produce none. Compounds with fewer than 32 hydrogen atoms were assigned to have post-cyclization decoration via double bond formation. More than H34/H36 atoms for macrocyclic and labdane-derived compounds respectively indicated additional post-cyclization saturation events. The relative frequency of H32, H34, H36, <H32, >H36, and aromatic ring systems for the top 20 most common diterpenes were all quantified (Figure 3.3).

Carbocations were estimated based on double bond positioned in conjunction with tertiary carbons or specifically the secondary carbon involved in kaurene synthesis [Tantillo 2010]. When a double bond was positioned at two neighboring tertiary carbons, both positions were considered as possibilities. In circumstances where carbocations were predicted to be quenched

with water, all variants with the predicted number of quenching events were created and compared to predict final carbocation position (i.e. 34H: One hydroxyl, 36H: Two hydroxyl groups). Compounds with a single unambiguous, resolved carbocation structure were estimated as final products. Compounds with multiple estimated structures were compared to the list of unambiguous structures. Those that were found to match a previously identified unambiguous structure, prioritized based on frequency, were quantified to that compound group. Compounds that were aromatic or did not match an unambiguous structure were identified as ambiguous and excluded among resolved carbocation structures (Figure 3.3; Supplemental Code 4).

#### ***Determining diterpene molecular activity from variability of atomic decoration and bonds***

Because of the higher compound count and inclusion of stereochemistry, the skeletons from the deconstructed TeroKit database (Supplemental Data 2b, 2f) were used to evaluate atomic variability within the top 20 most common diterpene classes (Figure 3.4; Supplemental Data 7). This analysis aligned all SMILES backbones with the same skeleton and compared each atom and bond location to the canonical reference (Supplemental Code 5). All backbones were harmonized to have the same number of non-hydrogen atoms. This configuration was solved again by replacing any carbons with R-groups instead with Xenon. If the canonical SMILES for compounds with 20 atoms completely aligned with the reference SMILES, the carbon stereocenters, “[C@]”, “[C@@]”, “[C@H]”, “[C@@H]”, were replaced with L, R, D, or U respectively. All occurrences of “[Xe]” were replaced with “X”. These conversions of stereocenters and “[Xe]” made it so all SMILE names were an identical number of atoms and characters L, R, U, and D were later converted back into “C” after the bonds with indicated stereochemistry were replaced with the symbols “^” and “\*” to represent the chirality of those

bonds. At completion this method generates an output similar to what is visualized below

(Supplemental Data 7):

```
C(-C)(-C)-C-C-C-C(-C)-C1-C-C-C(-C)-C2-C-C-C(-C)-C-C-1-2  
  
C(-C)(-C)=C-X-C-C(-X)-C1-C-C-C(-C)-C2:X:C:C(-C):X:C-1:2  
C(-C)(-C)=C-C-C-C(-C)-C1-C-C-C(-C)-C2:X:X:C(-C):C:C-1:2  
C(-C)(-C)=C-C-C-C(-C)-C1-C-C-C(-C)-C2:C:C:C(-C):C:C-1:2  
...  
...  
C(*C)(^X)-C-C-C-C(*C)^C1-C-C=C(-C)*C2-C-C-C(-C)=C^C-1-2
```

Occasionally a backbone SMILE did not conform to the canonical reference. To correct for this the canonical reference was iteratively changed to match the nonconforming SMILE by replacing bonds with the matching number of double bonds, triple bonds, or aromatic bonds equal to those present in the nonconforming SMILE. When checking for bonds, occurrences of Xenon were substituted with carbon. When the bond patterns matched, Xenon atoms iteratively replaced carbon in the updated compound until this also matched the non-conforming SMILE. In order to make sure all SMILES had conforming atom order to the canonical SMILE, two approaches were used. The first approach checked compound similarity as each bond or Xenon was replaced, then used the top 5 most similar compounds to the product at each step until number of bonds and Xenon matched the non-conforming SMILE (comparing <60 compounds per iteration). When this method did not work the second approach was implemented. This approach substituted the number of bonds and Xenon from the nonconforming SMILE in every possible combination in reference to the canonical SMILE. This has the potential to generate many compounds for comparison and directly parallels the binomial coefficient equation to estimate compound abundance. For example, if a compound contained ten Xenon atoms, there would be ~184,000 different possible compounds, and only one will match our nonconforming

reference in canonical order. This method, while highly accurate, was only used as a last resort and faces even greater computational demand when working with molecules larger than diterpenes.

When backbones aligned, the reference SMILE was converted to an edge/node table (Supplementary Code 5). The level of diversity for each bond (edge) and atom (node) is calculated using the index of qualitative variation (IQV). All occurrences of R-groups were assumed to be unique. IQV values for atoms were used and saved as intensity values for each node and the IQV values for bonds were used to describe the intensity of each connecting edge. Edge/node tables were used as input Cytoscape (v3.10.1) to visualize the data (Figure 3.4; Supplemental Data 7) [Shannon et al. 2003].

Decoration bias based on carbon position (primary, secondary, or tertiary) for the two major classes, labdane-derived and macrocyclic-derived, were investigated. This was accomplished by calculating the percentage of decoration at each carbon location with consideration to the number of connecting carbons and the position of all neighboring carbons (Supplemental Code 5). All conditions were then compared to determine effects based on diterpene class and position-based variation. Only Carbon positions that had at least three instances of atomic location and neighbors among labdane- or macrocyclic-derived diterpenes were used.

### ***Phylogenetic distribution of diterpenes within Viridiplantae, Rhodophyta, and Chromista***

The DNP skeleton summary datasets (Supplemental Data 2a, 2e) were used to quantify compound abundance among land plant families, green algae, red algae, and brown algae (out group) (Supplemental Code 6). Plant taxonomic Family was extrapolated to also consider a Phylum and Kingdom level analysis. An indexed list was created matching the reported Family to compound. This list was compared to the diterpene skeleton list (Supplemental Data 2e) to

identify overlap and quantify reported skeletons for all selected Families. Families were excluded from phylogeny if they had less than 10 reported compounds among the top 50 most common diterpene skeletons. The remaining Families were manually sorted based on phylum divergence and top diterpene skeletons were manually annotated based on DNP reports and external reports [Eman et al. 2020; Feng et al. 2014; Gao et al. 2016; He et al. 2005; Li et al. 2013; Mendes et al. 2023].

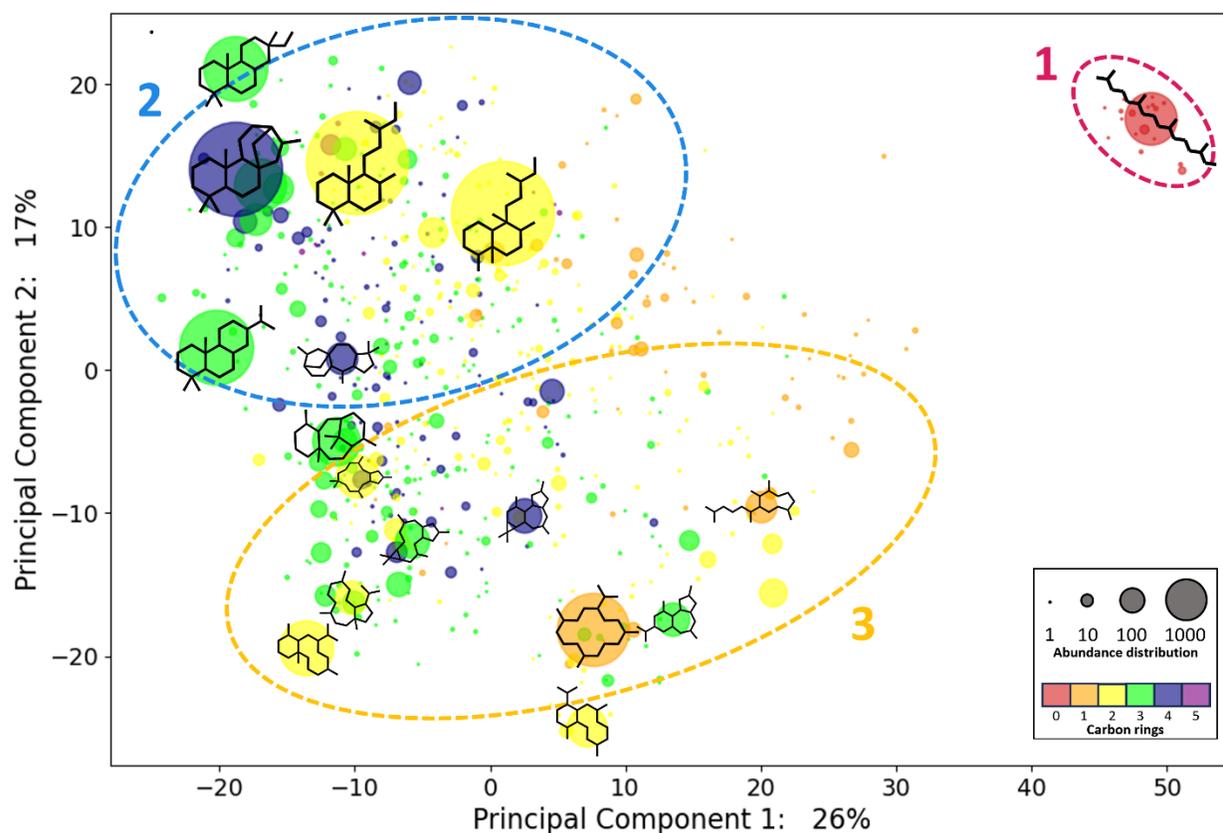
## **Results & Discussion**

### ***Diterpene database deconstruction summary***

Among the reported diterpenes (>60K) we identified a total of 924 unique C<sub>20</sub> skeletons identified from the TeroKit (872) and DNP datasets (671) (Supplemental Code 1). Only C<sub>20</sub> skeletons were considered as skeletons candidates because nearly all structures that were not C<sub>20</sub> were either misannotated, derivatives of identified C<sub>20</sub> diterpene structures, or diterpene-diterpene dimers (40C). Additionally, the majority of both databases could be deconstructed back to C<sub>20</sub> (TeroKit 85.8%; DNP 88.0%). Among the non- C<sub>20</sub> skeletons identified in TeroKit, backbone with C<sub>19</sub> (2960), C<sub>18</sub> (691), C<sub>17</sub> (316), and C<sub>21</sub> (312), mainly due to (de)methylation events of already identified C<sub>20</sub> backbones. The majority of diterpenes were represented by a small percentage of structures, with the top 25 most common skeletons defining over 75% (26,000 compounds; Supplemental Data 2g). Echoed in the findings of previous work [Johnson et al. 2019a], monoterpenes (~60), sesquiterpenes (~320), triterpenes (~70), and other compounds (such as alkaloids) were misannotated as diterpenes (Supplemental Data 2h; Supplemental Code 1). While the compound landscape is filled with anomalies, being able to categorically organize all reports allows for the parsing and curation of the data as a whole,

which can determine whether a compound is or is not a diterpene. This has the potential to provide a select group compounds for manual database filtration, with the potential of removing near all misannotated compounds.

Skeleton structures from the DNP were compared to visualize compound distribution and similarities (Figure 3.1). Skeletons clustered largely based on mode of cyclization, representing 26% of identified skeleton variance, where the linear phytane backbone and other minor acyclic derivatives, not acted upon by TPSs, distinguished themselves from those that were cyclized (Figure 3.1, Group 1). The next major clusters identified 17% of variance and were represented by the labdane-derived structures (Figure 3.1, Group 2), defined by their synthesis via diTPS class II/class I activity compared to the macrocyclic-derived structures (Figure 3.1, group 3), which are typically acted upon by a class I diTPS only.



**Figure 3.1:** Principal component analysis of DNP diterpene skeleton structures based on RDKit bit vector comparison scores

Each point represents a diterpene skeleton in which point size is correlated to the number of reported compounds in the DNP database that deconstructed back to that specific core structure. Point color indicates the number of carbon rings present for each structure. PC1 defines 26% variance and is largely characterized by the presence or absence of rings. PC2 defines 17% variance and distinguishes itself based on mode of TPS cyclization. Cluster 1 represents the phytanes, which are the dephosphorylated form GGDP, the most common diterpene precursor. Cluster 2 contained the labdane-derived class II/class I compounds. Cluster 3 contained macrocyclic-derived compounds, synthesized via class I activity.

### *Modeling diterpene backbone cyclization validates known and identifies new chemistries*

We modelled diTPS synthesis to validate the origin of observed skeleton diversity and investigate additional theoretical chemistries. Performing reactions in two steps allowed chemistry to mirror the two main methods of diTPS activity, either through class II mechanism or class I mechanism. The minimum number of steps required to produce clerodane (class

II/class I product) and dolabadiene (class I product) each required 8 generations. The increase to 10 generations for both class II and class I reactions generated only 3 additional skeletons, indicating approach towards the upper threshold. Quantification of class II and class I products generated >347,000 compounds and matched to a total of 80 unique skeleton structures (Figure 3.2A).

The acquisition and implementation of the computational rules varied greatly within the datasets. There were 12,319 occurrences where multiple rules created the same product. While some rules were intended to perform very specific and limited reactions, in practice this was not always the case as structures continued to become more complex. For example, amphilectane biosynthesis can only occur in one way and as predicted the necessary reaction rule only met conditions for biosynthesis 6 times. This contrasted with a different rule that was designed with precise chemistry, which forms the secondary beyeranyl carbocation necessary for pimarane to cyclize to beyerene, kaurene, artisinene, and trachylobane [Hong & Tantillo 2011]. In practice, this very specific secondary carbocation rule met conditions for implementation over 1,000 times. Rules designed to be more generic on the other hand, such as those to quench carbocations, hydride shifts, or methyl shifts understandably occurred >254,000, >115,000, and > 24,000 times respectively. As would be expected, these include rules for the quenching of carbocations, which can occur at every step in the process. Likewise, 1,2-hydride and 1,2-methyl shifts, which only required an adjacent tertiary or quaternary Carbon respectively and become more common as structures are more cyclized.

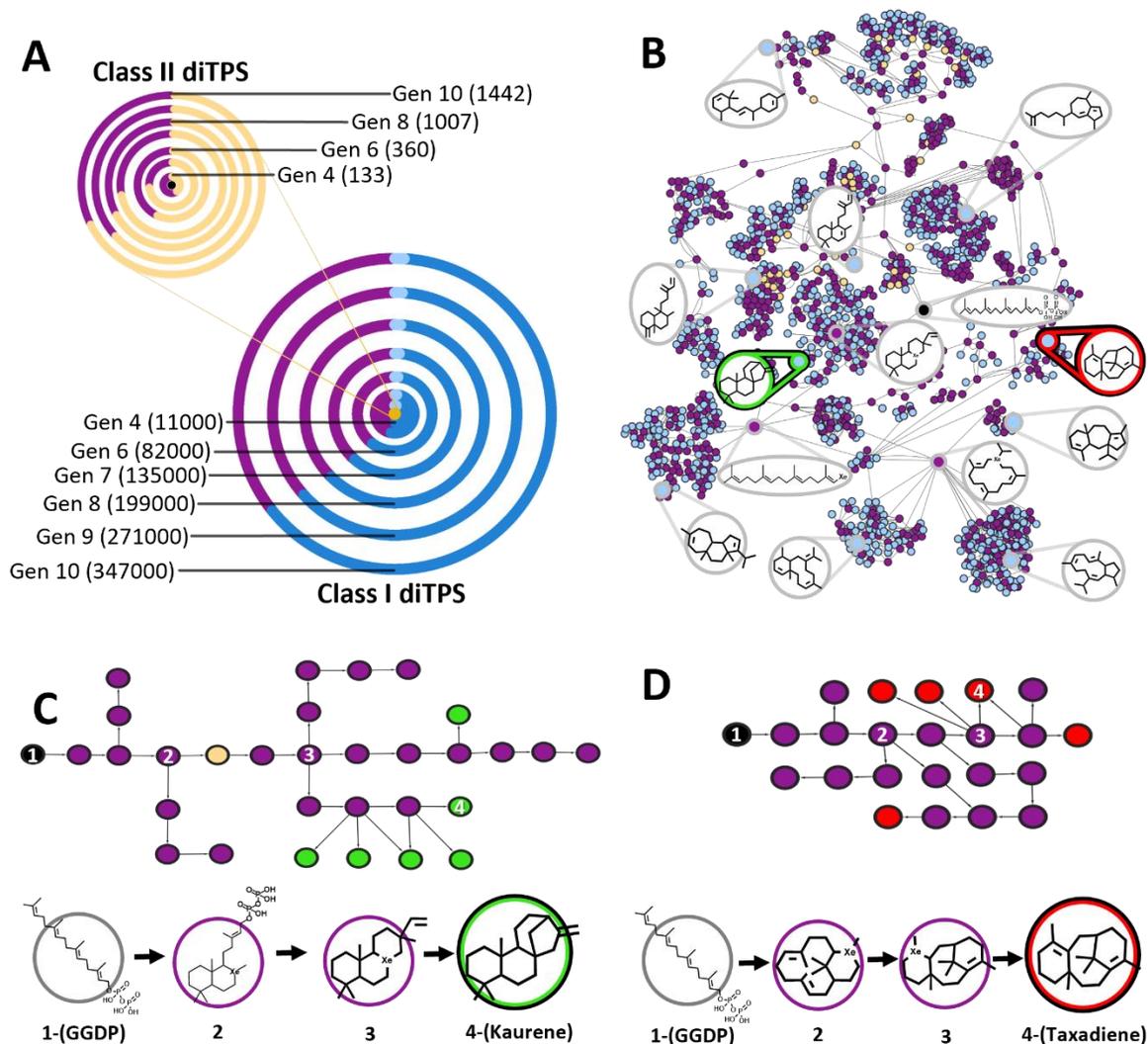
In total, actions orchestrated by diTPSs exclusively are able to determine 80 of the identified skeleton structures. Notable compounds not represented among the synthesized skeletons included the lathyranes, ingenanes, tiglianens, and grayananes. The lathyrane and lathyrane-

derived skeletons are specifically known to originate through initial alcohol dehydrogenase activity and thus were also expected to be absent here [Luo et al. 2016; Wong et al. 2018]. Grayananes are speculated to be an oxidative rearrangement derivative from *ent*-kaurene, with synthesis exclusive to the Ericaceae family [Fay et al. 2022; Turlik et al. 2019]. As one of the more prominent compounds in the database, grayanane biosynthesis illustrates an interesting phenomenon shown within the model. While only 80 of the 924 skeletons were able to be recreated solely through diTPS activity, much of that remaining diversity (like is the case with grayanane) still relies on diTPS driven carbocation rearrangement for synthesis. Then, after cyclization, many other, less common enzymatic processes provide mechanisms beyond the scope of diTPS and carbocation activity leading to additional skeleton diversity.

While only 10% of skeleton diversity was represented by unique diTPS carbocation rearrangements, many structures require additional secondary derivatization from the actions of other non-diterpene synthase enzymes, for example, the structures seen with the grayananes, gibberellins, and seco-kaurenes [Fernández-Martín et al. 1995, Turlik et al. 2019, Fay et al. 2022, Zou et al. 2023]. When rules are applied to change skeleton shape using rules beyond the scope of diterpene activity, 608 additional skeletons origins were identified. These alternative rules provided the additional diversity through the action of ring breakages (86), alternative ring formation (59), carbon side chains shifting (173), ring and methyl groups collapsing/expanding (35), or a combination of these conditions (255). The remaining ~240 skeletons that were unaccounted for likely require highly specialized mechanisms, are plausibly derived from an alternative C<sub>20</sub> prenyl-diphosphate [Cheng et al. 2012, Miller et al. 2020] or may have been misannotated as diterpenes altogether.

A total of 737 unique compounds represented the 80 diTPS derived skeleton and these were used to generate a reaction network and predict origin of backbone synthesis (Figure 3.2B, 3.2C, 3.2D). Cembrene, pimarane, and taxane related carbocations represented the highest network stress and betweenness centrality, which are indicative of nodes that have high flow through or serve as a critical hub required for further synthesis [Shannon et al. 2003]. Two Cembrenyl ion variants had the highest significant representation as a critical hub, having the first and seventh highest betweenness centralities of 0.372 and 0.249 respectively. The first ion, the main cembrenyl cation paralleled previous reaction reports leading to direct synthesis of cembrene A and cembrene C [Meguro et al. 2013; Rinkel et al. 2018]. The cembrenyl cation also connects downstream with a wide variety of macrocyclic diterpenes like taxanes and casbanes, which serve as additional relevant precursors to greater skeleton diversity among the Taxaceae and Euphorbiaceae families respectively (Figure 3.5) [Luo et al. 2016; Rinkel et al 2017; Li & Dickschat 2022]. This ion variant had a hydride shift post-cyclization, indicative of other complex reactions. Some of the reactions available with this ion, like apparent secondary carbocations or ranged proton transfers, exemplify instances of how hyperconjugation or charge delocalization originating from different parts of the structure can permit otherwise unlikely reactions. The nature of investigating these mechanisms in a stepwise fashion makes some of these instances artifacts of the model, while still highlighting the pivotal roles played based on certain atomic configurations [Hong & Tantillo 2011; Tantillo 2011; Schrepfer et al. 2020]. Coincidentally, this variant provides an important first step in Pickaxe to create some of the most common reported marine-life derived skeletons, such as the briaranes and eunicellanes [Moon & Harned 2018; Rinkel et al. 2018; Xu et al. 2022; Yan et al. 2023]. It is of note that the eunicellanes are synthesized differently in coral and bacteria and both mechanisms are

represented here, but the mechanism implemented by coral is caused by this particular cembrenyl cation [Li et al. 2023; Li & Rudolf 2023]. Likewise, because the pimaranyl cation (0.345 betweenness centrality) is necessary to form trachylobane, abietane, atisane, beyerene, kaurene, (iso)pimarane, cassanes, and more, it is reasonable that it had the second highest betweenness centrality [Tantillo 2010]. Lastly, a verticillanyl cation (precursor to taxanes and abeotaxanes) and a taxadiene ion also saw particularly notable betweenness centrality (0.321 and 0.284 respectively).



**Figure 3.2:** Summary of carbocation cyclization (TPS enzyme) reactions modeled at a global/theoretical level, filtered to identified structures, and examples of local synthesis.

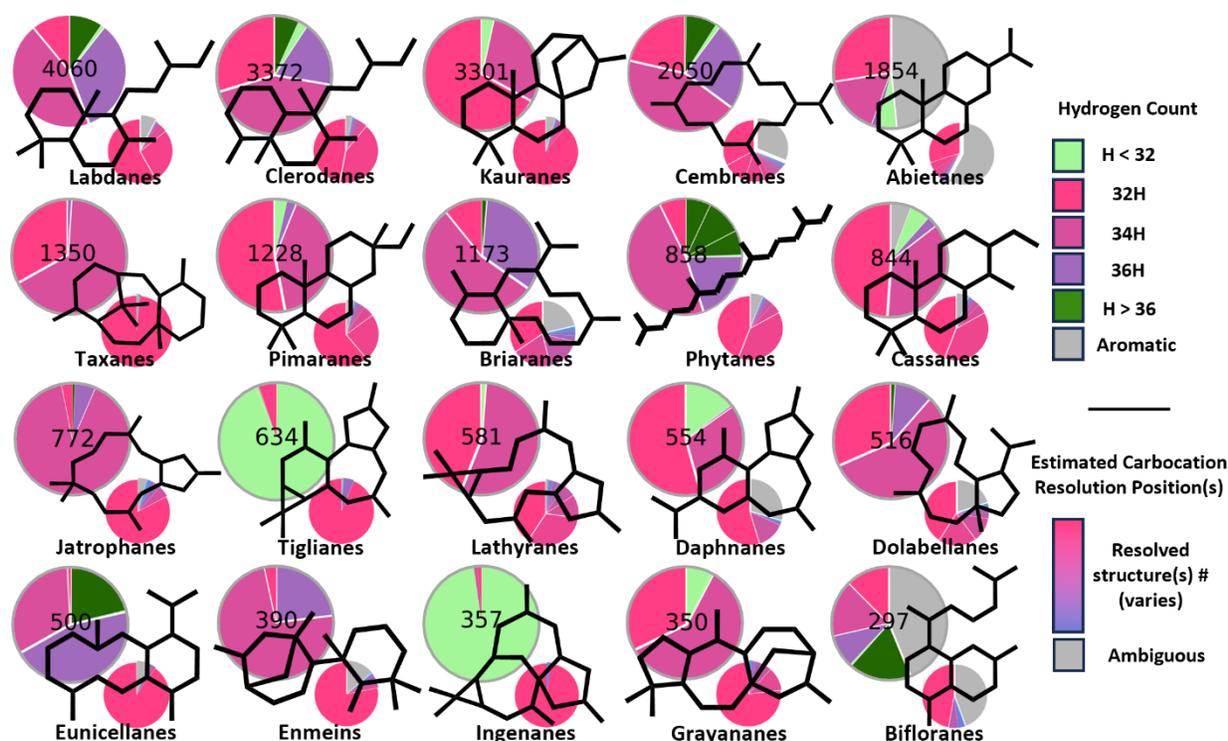
Structures throughout are indicated as either **Black** (GGDP), **Purple** (carbocation containing), **Yellow** (phosphate retaining), **Blue** (resolved backbone), **light Blue** (resolved backbone matching a known structure), **Green** (kaurene), or **Red** (taxadiene). **A** Products of diTPS class II reactions and created 1442 total structures after 10 generations. The second iteration built upon the first where resolved compounds with retained diphosphate (931 structures (yellow) and GGDP (black)) were used as inputs. The second iteration performed diTPS class I reactions, creating 347,000 compounds after 10 generations. **B** The third Pickaxe iteration filtered to reaction paths exclusively to those that led to known structures. Local examples of synthesis are visualized for **C** the class II/class I synthesis of kaurene, and **D** the class I synthesis of taxadiene.

### *Predictive carbocation quenching patterns*

Carbocation cycloisomerization prevalently generates tertiary carbocations as they tend to be the lowest energy state intermediates [Peters 2010; Tantillo 2010]. Beyond the cycloisomerization, carbocations are generally resolved either by a deprotonation to form a double bond or through quenching with water. This leads to post-cyclized backbones having a mass of 272 g/mol, 290 g/mol, or 308 g/mol, depending on whether zero, one, or two carbocations were resolved with water. In the process of synthesis, labdane-derived compounds generate two carbocations, macrocyclic-derived compounds generate one carbocation, and phytanes generate zero. Because of this phenomenon, the carbocation resolution methods, additional double bond (de)saturation events, and aromatization events can be estimated based on final hydrogen number (Figure 3.3). Generally, each compound class illustrates unique carbocation resolution and post-cyclization modification patterns, with closely related structures sharing some similarities. For example, the labdanes and clerodanes share remarkably similar quenching and post-cyclization modification distribution, however that distribution differs greatly from the kauranes and abietanes. Only three skeleton classes had reported aromaticity among their compounds, with those being the abietanes, cassanes, and bifloranes. Because of their absence among the other compound classes, it appears a slight majority of reported abietanes and bifloranes were aromatic (~50%; Figure 3.3). The bifloranes and abietanes are both known to have spontaneous aromatization events. For the abietanes, miltiradiene is known to spontaneously aromatize to form abietatriene [González 2015; Zi & Peters 2013; Bryson et al. 2023]. Likewise, the biflorane derivative, dihydroserrulatene is known to spontaneously aromatize to serrulatane [Zi & Peters 2013; Miller et al. 2020].

Certain skeleton classes had the majority of molecular features represented by one predicted orientation, particularly in the case of the kauranes, enmeins, grayananes, taxanes, eunicellanes, and tiglianes (Figure 3.3). These classes all had carbocations that could be resolved back to a single method of carbocation quenching among reported compounds. This could be because of a necessity for conservation due to essential function, as is the case of the kauranes as hormone precursors. Alternatively, commonality in skeleton reports may be shared because a backbone is isolated to a particular species or sampling bias due to interesting bioactivity, which may be the case for skeleton classes such as the taxanes, tiglianes, and ingenanes. Because enmeins and grayananes are likely derivatives of kaurane, it is of note that the core quenching patterns seem to remain similar in these structures after the major skeletal modifications take place (Figure 3.3). This phenomenon is particularly curious for the eunicellanes due to the aforementioned two diverging mechanisms leading to the same product [Li et al. 2023; Li & Rudolf 2023].

The majority of phytane derivatives have double bond modification, as ~90% of compounds have fewer double bonds than GGDP. This is the case for many macrocyclic-derived compounds as well. Extrapolating from this, the macrocyclic-derived cembranes, briaranes, and eunicellanes all see double bond desaturation as common post cyclization modification (represented by purple and forest green; Figure 3.3). Alternatively, abietanes, cassanes, daphnanes, and especially the tiglianes and ingenanes see desaturation as a common post-cyclization decoration (represented by mint green; Figure 3.3).



**Figure 3.3:** Summary of carbocation quenching patterns and post cyclization decoration for each of the top 20 most common diterpene skeleton classes in the TeroKit database.

Deconstructed diterpene backbones for each skeleton class were investigated based on carbocation resolution patterns. Resolutions occurred via rearrangement or quenching with water, which generally led to a final hydrogen count of 32, 34, or 36 (**Pink, Magenta, Lavender** respectively). Occurrences of less than 32 hydrogens suggest additional backbone modification after cyclization, leading to double bond formation (**mint Green**). The occurrence of more than 36 hydrogens suggest double bond desaturation modification, (**forest Green**). Indicated by the smaller pie chart are the quantified estimates of unambiguous carbocation structures. Aromatics and ambiguous structures were identified with **Gray**.

### ***Multiple SMILES alignment of diterpene backbones identifies atomic “hotspots”***

Localized atom and bond diversity within specific diterpene classes was investigated using IQV values for each atom and bond among the top 20 TeroKit diterpene skeletons (Figure 3.4). This was done to speculate the effects of position-based sources of bioactivity, evolutionary driven effects, steric availability for decoration, and/or localized rigidity for common diterpene cores. Notably, the labdane-derived and macrocyclic-derived compounds show distinctions in

decoration (Figure 3.4A). Regardless of neighbor or position, labdane-derived compound carbons are decorated at a rate of ~20.5% (Supplemental Code 5). One commonality to functional group addition being that the primary carbon, which neighbored the diphosphate prior to cyclization, was often decorated at a frequency of ~56.2% (Figure 3.4B). Another identified commonality being that one of the two methyl groups that neighbor each other is often more decorated. This localized decoration is likely a product of the CYP701 family, which is specifically known to target this region and its ability to act on a range products may be due to neofunctionalization within the enzyme family or decoration occurring early in labdane formation [Bak et al. 2011].

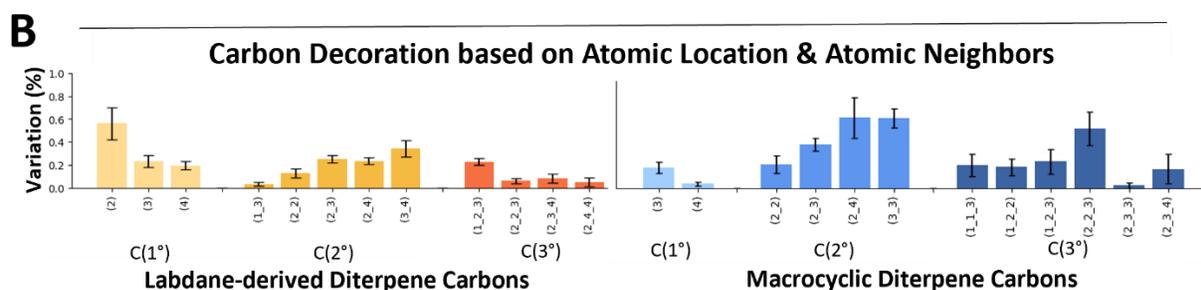
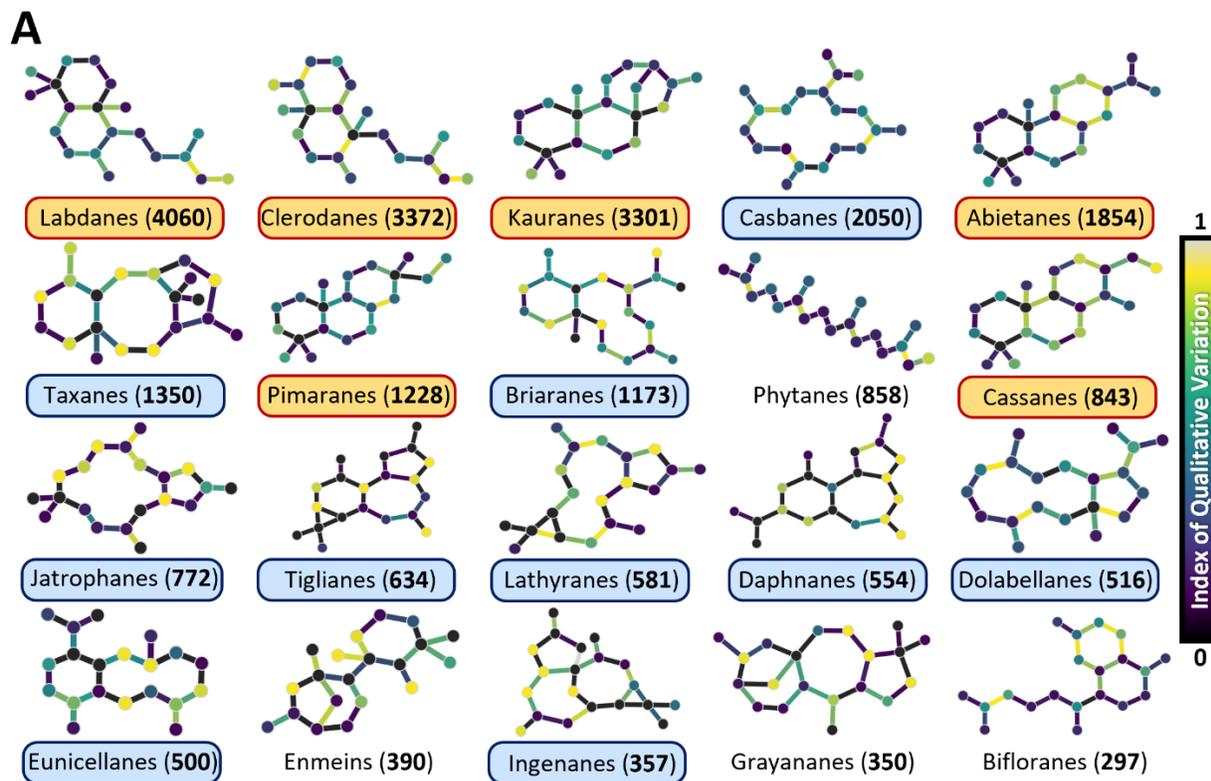
Alternatively, the macrocyclic-derived compounds displayed much higher atom variation overall, with a decoration rate of ~28.7%. Among bond specific variations, the tertiary carbons with a methyl group neighbor, saw the highest fluctuation in bond type (Figure 3.4A). Additionally, many of the macrocyclic-derived compounds had exceptionally high frequency of atom decoration. Of note, some macrocyclic derived atomic positions had decoration for nearly every reported backbone (>95% decoration rate, Supplemental Data 7). Generally, these compounds saw the highest decoration among secondary carbons, with decoration occurring ~52.9% of the time (Figure 3.4B). This high frequency may be partially explained by some macrocyclic compound originating from one pathway but with promiscuous timing of when certain oxidative decorations are added.

Overall, kauranes had the lowest decoration overall, despite having over 3,300 unique entries. Due to its main role as a plant hormone precursor, this limited diversity in structure may be necessary to retain its integral role in signaling, compared to many other groups, which more commonly function in host defense. Interestingly however, the enmeins and grayananes, which

are speculated to be derivatives of kaurene [Ujita 1972; Yang et al. 2016; Pan et al. 2018; Turlik et al. 2019; Fay et al. 2022], have atom/bond variation that is more similar to the macrocyclic-derived compounds than kauranes. The enmeins are mainly sourced from Lamiaceae, which also host the highest abundance of kaurane entries as well (Figure 3.5) [Zeng et al. 2020].

Grayananes on the other hand are nearly all sourced from the Family Ericaceae (Figure 3.5), and with this high frequency of decoration may be specific just to that Family. This large divergence in atomic decoration may suggest that skeleton distance in grayananes and enmeins can only occur after substantial structural derivatization from kaurane has taken place first.

Taxanes, jatrophone and derivatives (ingenanes and lathyranes) also provide a unique representation among the dataset. Atoms of these compounds represent atomic diversity where each location seemed to have an all-or-nothing representation, where atoms were nearly always or almost never decorated (Figure 3.4A). Because these compounds are known to have high bioactivity [Guenard et al. 1993; Croteau et al. 2006; Luo et al. 2016], this distinction among compound classes may be a product of their utility in nature, leading to the diversification and expansion of compounds with high bioactivity within the Euphorbiaceae and Taxaceae Families and only building upon successful diterpene derivatives. Alternatively, this phenomenon may be due to sampling bias and growing interest for specific sources and compounds in the context of research pursuits within these families. For example, bioactive compounds, such as Taxol [Guenard et al. 1993; Croteau et al. 2006], may be used and derivatized more often because of their importance in pharmaceuticals applications, leading to specific compound classes seeing a more “all-or-nothing” response within the database.



**Figure 3.4:** Visual of atom and bond variation among the top 20 most common diterpene skeletons and identified variation related to diterpene origin, carbon connection(s), and carbon neighbor(s).

Skeleton names and the number of compounds with that representative skeleton were labeled for each. Compounds were distinguished as either labdane-derived (**Orange**), Macrocyclic (**Blue**), or ambiguous to these distinctions (**White**). Atom and bond variation was calculated using the index of qualitative variation (IQV). Variation of carbon position (1°, 2°, or 3°) and the position of all neighbors were investigated to identify positional based decoration for both Labdane-derived and Macrocyclic compounds.

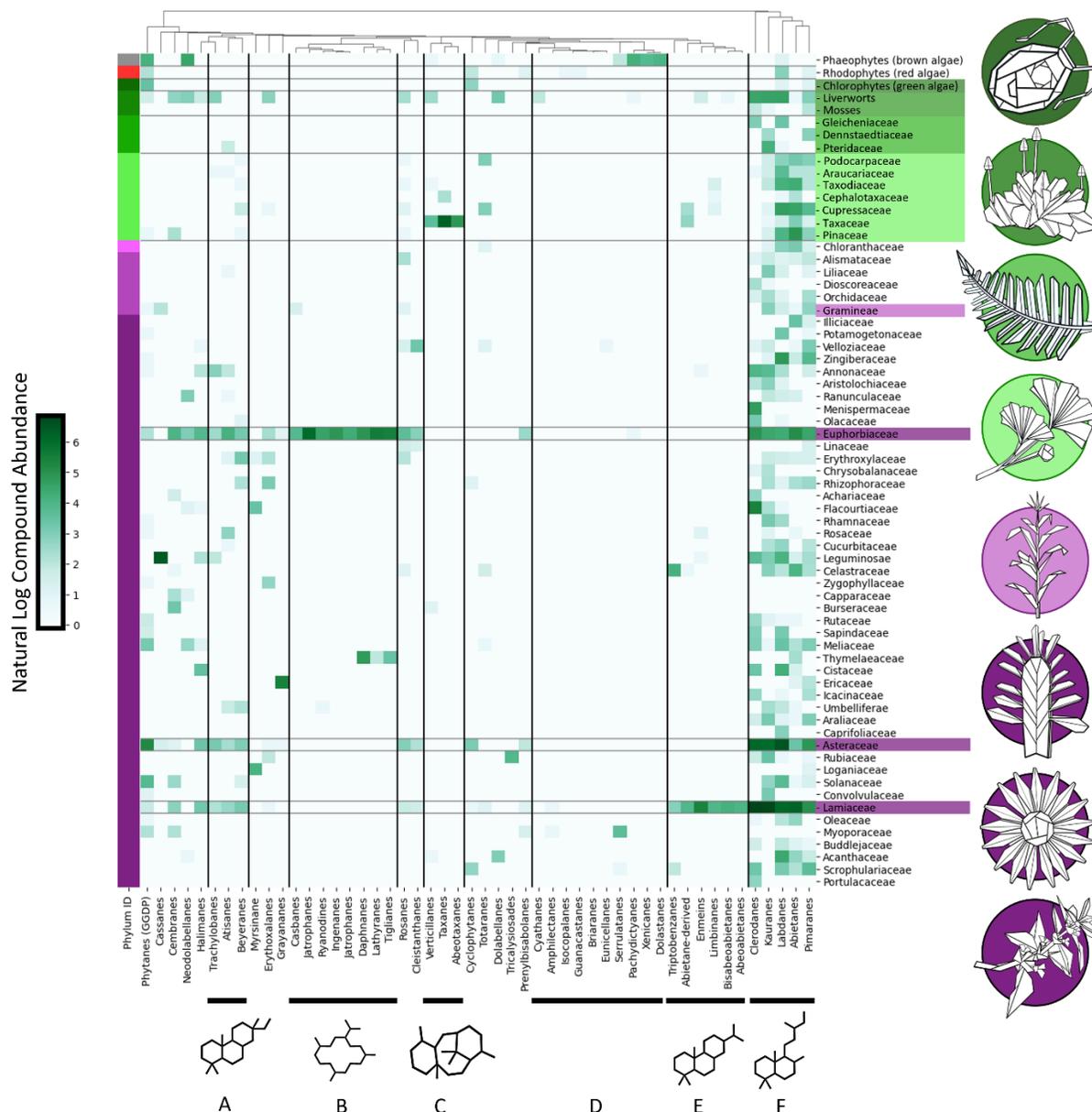
***Distribution and magnitude of diterpene diversity in relation to evolutionary divergence among Viridiplantae, Rhodophyta, and Phaeophytes***

Plants are the main known producer of diterpenes, which is echoed in the DNP database (Supplemental Data 1a, 2a) and in previous reports [Zeng et al. 2020; Johnson et al. 2019a]. There are a total of 196 unique taxonomic Families in the DNP and 149 represent Plantae or algae. Paralleling this 76.9% (16761/21804) of all database compounds also come from plants. Specialized metabolism is often considered to be a potential driving force in speciation [Ehrlich & Raven 1964; Pichersky & Lewinsohn 2011; Chae et al. 2014; Alicandri et al. 2020; Jia et al. 2022] and here it was investigated to identify if certain structures and their presence or absence correlated with phylogeny at a family level. The top 50 DNP skeletons (Supplemental Data 2a, 2e), representing 84.5% of all DNP diterpene entries, were used as the first metric, and the 67 plant Families, Charophytes (green algae), Rhodophytes (red algae), and Phaeophytes (brown algae) with at least 10 diterpenes reported were compared (Figure 3.5). Overall little correlation between reported diterpenes and phylogenetic divergence at a Family level were observed, with the exception that most Gymnosperms (lacking Taxaceae and Cephalotaxaceae) clustered together (Supplemental Data 8a). This does not discount that speciation may still occur at a more local level, but the presence and absence of diterpene skeleton derivatives appear to be more dynamic in their presence and origin (Figure 3.5).

Regarding diterpene abundance, the Lamiaceae (3,595 diterpenes reported), Euphorbiaceae (2,285 diterpene reported), and Asteraceae (2,073 diterpenes reported) had the highest respective compound entries (Figure 3.5). These families are also some of the largest and most diverse plant clades (Asteraceae ~24,000 species; Euphorbiaceae ~7,000 species; Lamiaceae ~7,000 species), but no significant correlation was identified between the number of total species within a family and compound abundance (Supplemental Data 8b, Supplemental Code 6). Also of note, the

Taxaceae (27.93), Ginkgoaceae (10.00), Taxodiaceae (9.05), and Cephalotaxaceae (6.40) had the highest ratios of reported diterpenes per species, corrected for family size (Supplemental Data 8c, Supplemental Code 6).

Comparatively, skeleton distribution and abundance clustered distinctly due to a number of factors. The first major cluster (A; Figure 3.5), represents further derivatized versions of the pimarane skeletons, with them largely present in Liverworts and Angiosperms. Next, the Euphorbiaceae dominated reports of macrocyclic structures (B; Figure 3.5), and represented compounds further derived from the casbane and cembrane skeletons. The following cluster (C; Figure 3.5) contains compounds tied closely to the Taxaceae family, representing taxane and taxane-derived skeletons. Major compounds most commonly synthesized in fungi, coral, and bacteria (D; Figure 3.5) are prominently absent among most plant Families, with some unique skeletons only reported within the Phaeophytes (brown algae), which diverged from Viridiplantae over 1 billion years ago [Yoon et al. 2009]. The next major cluster contained abietane derivates (E; Figure 3.5), which were nearly exclusive to Lamiaceae. The last major cluster (F; Figure 3.5) contained compounds widely distributed among all Plantae (8,725 compounds; 39.6%) and represent the major labdane derivatives, including the clerodanes, abietanes, pimaranes, and kauranes.



**Figure 3.5:** Heatmap of top 50 most common diterpene skeletons and their abundance in plants and algae.

Families were manually organized based on Phylum divergence in respect to each other (**Grey:** brown algae, **Red:** red algae, **Green (dark to light):** Charophytes, Bryophytes, Lycophytes, and Gymnosperms, **Purple (light to dark):** Magnoliids, Monocots, Eudicots). Compounds were hierarchically clustered and were distributed based on **A** derived skeletons from pimaranes, **B** cembrene and casbene derived diterpene skeletons in Euphorbeaceae, **C** taxane and taxane derivatives, **D** diterpene skeletons most commonly found in fungi and marine life, **E** derived skeletons further derived from abietanes, and **F** Labdane-derived compounds.

## Conclusion

The Dictionary of Natural Products and TeroKit databases provide an important and monumental source of compound information, however the extraction of useful or curated data can be difficult. Work presented here improves upon previous curations of diterpene libraries, provides improved method for identifying, dissecting, and synthesizing diterpenes, along with predicting different biases and errors present among current database submissions [Johnson et al 2019a; Zeng et al. 2020; Zeng et al. 2022] (Supplemental Data 2a-2f). Paralleling previous findings, the majority of reported diterpenes are mainly sourced from Euphorbeaceae, Lamiaceae, Asteraceae, and multiple marine-life sources [Johnson et al 2019a; Zeng et al. 2020; Zeng et al. 2022]. Additionally, the overall diterpene landscape is represented by a limited number of skeletons, with 25 of the 924 identified structures representing over 75% of the complete database [Johnson et al. 2019a, Zeng et al. 2020]. While most diterpene skeletons begin either through the class II/class I or class I diTPS mechanisms, only 80 diterpene skeletons could be produced exclusively through these means. While more terpene synthase mechanisms likely exist, this select group is indicative that the majority of terpene diversity instead comes from additional activity. This trend is further emphasized by the additional 608 skeletons that are somehow derived from more unconventional modifications to common diterpene skeletons by means of ring breakages, shifting carbon chains, or other sources.

The dissection and construction of the diterpene library in this way has illuminated a number of notable patterns as well. This includes differences between the labdane-derived and macrocyclic-derived compounds in their reported decoration patterns. The kauranes and taxanes both act as unique outliers within the dataset, for different reasons. Kaurane is the third most abundant reported diterpene skeleton and has many post TPS modified skeletons as well (including the

enmeins, grayananes, gibberellins, and so on), all of which generally have high degrees of variability in their decoration and core structure. Despite kaurane having over 3,000 unique entries, is one of the most conserved structures with overall limited decoration and structural variance among any of the common skeletons (Figure 3.3; Figure 3.4). Taxanes represent a family of compounds with little variation in core structure formation (Figure 3.3) but see high conservation for atoms which either are or are not decorated among reported samples (Figure 3.4). Because of the high interest and bioactivity of taxadiene derivatives in medicine [Wani et al. 1971; Croteau et al. 1993], semi-synthetic derivatives [McGuire et al. 1997; Hao 2021], and generally conserved structure, the taxanes are likely reported highly not because of the many derivatives present in nature (as only few species produce them) but instead because of sampling biases, and semi-synthetic derivatives. Also, the high ratio of reported compounds to family size among Taxaceae and taxanes may suggest an error in the reported sourcing among some taxanes as well (Supplemental Data 8c; Figure 3.5).

The software used to deconstruct the diterpene library (Supplemental Code 1) and reconstruct diTPS synthesis mechanisms (Supplemental Code 3a) both provide modular platforms for accommodating additional considerations, a growing database size, and revision in the future. These tools in conjunction with each other provide an excellent framework for modelling and predicting the mechanistic modes representing the current diterpene landscape. Formatting many of these scripts in Jupyter Notebook also provides an easy-to-pickup tool for those with more limited experience in Python [Kluyver et al. 2016; Ferretti et al. 2019; Kumar et al. 2023]. Pickaxes' plug-and-play nature also allows individuals to pick up presented rulesets, modify, and expand upon them. The TeroKit and DNP datasets as they currently stand are excellently curated and have an incredible wealth of knowledge. This exploration has enlightened that they likely

have an error rate of less than 1% however the complete and automated deconstruction of the library, done here, provides avenues for improving curation of these databases to near perfection. In future versions of this work, these models could be improved further with the inclusion of additional diTPS rules, writing prevalent reaction rules beyond diTPS activity exclusively, considering stereochemistry, and applying thermodynamic parameters without penalizing concerted reaction intermediates.

### Data availability

The following supplemental data have been deposited at:

<https://doi.org/10.5061/dryad.ksn02v7cb>

Supp.1a.DNP\_Diterpene\_Mining\_v30.2.csv

Supp.1b.TeroKit\_Diterpene\_v2.0.tsv

Supp.1c.TeroKit\_Triterpene\_v2.0.tsv

Supp.1d.TeroKit\_Sesquiterpene\_v2.0.tsv

Supp.2a.DNP\_Diterpene\_Mining\_v30.2\_Updated\_Skeleton\_Backbone.tsv

Supp.2b.TeroKit\_Diterpene\_v2.0\_Updated\_Skeleton\_Backbone.tsv

Supp.2c.TeroKit\_Triterpene\_v2.0\_Updated\_Skeleton\_Backbone.tsv

Supp.2d.TeroKit\_Sesquiterpene\_v2.0\_Updated\_Skeleton\_Backbone.tsv

Supp.2e.DNP\_Diterpene\_Skeleton\_Summary.tsv

Supp.2f.Terokit\_Diterpene\_Skeleton\_Summary.tsv

Supp.2g.Terokit\_Diterpene\_Skeleton\_Abundance\_Distribution.png

Supp.2h.Terokit\_Diterpene\_Deconstructed\_Skeleton\_Carbon\_Distribution.png

Supp.3.PCA\_comparison\_matrix.tsv

Supp.4.NICKS\_FOLDER\_OF\_CARBOCATION\_EXPLANATION

Supp.5a.step1\_10\_gen\_Class\_II

Starting\_reactants.csv

classII\_diTPS\_rules.tsv

metacyc\_coreactants\_v2.tsv

step1\_10\_gen\_Class\_II\_reactions.tsv

step1\_10\_gen\_Class\_II\_compounds.tsv

Supp.5b.step2\_10\_gen\_Class\_I

Starting\_reactants.csv

classI\_diTPS\_rules.tsv

metacyc\_coreactants\_v2.tsv

step2\_10\_gen\_Class\_I\_reactions.tsv

step2\_10\_gen\_Class\_I\_compounds.tsv

Targets\_hit.tsv

Rules\_Ratios.tsv

Supp.5c.step3\_Real\_Backbones\_Only

-starting\_reactants.tsv

- confirmed\_backbone\_targets.tsv
- metacyc\_coreactants\_v2.tsv
- classII\_diTPS\_rules.tsv
- step3\_Class\_II\_reaction.tsv
- step3\_Class\_II\_compounds.tsv
- Supp.5d.non\_diTPS\_ruleset\_exploration
  - diTPS\_Skeleton\_reactants.tsv
  - non\_diTPS\_ruleset.tsv
  - metacyc\_coreactants\_v2.tsv
  - non\_diTPS\_reactions.tsv
  - non\_diTPS\_compounds.tsv
- Supp.6.Cytoscape\_Networks
  - Pruned\_Network (Figure 2b)
  - Clustered Network for Class I
  - Kaurene Network (Figure 2c)
  - Macrocylic Network (Figure 2d)
- Supp.7.Backbone\_IQV.tar.gz
- Supp.8a.Family\_Compound\_Phylogeny\_Clustered.png
- Supp.8b.Family\_Species\_count\_V\_compound.png
- Supp.8c.Family\_Species\_count\_V\_compound\_ratio.png

The following supplemental code has been deposited at

<https://doi.org/10.5061/dryad.ksn02v7cb>

- Supp\_Code.1.Terpenoid\_Deconstruction.ipynb
- Supp\_Code.2.Skeleton\_PCA.ipynb
- Supp\_Code.3a.Pickaxe\_DM\_NS.py
- Supp\_Code.3b.Match.py
- Supp\_Code.3c.Network\_maker.py
- Supp\_Code.4.Carbocation\_Quench\_Predictor.ipynb
- Supp\_Code.5.Backbone\_MSA.ipynb
- Supp\_Code.6.Phylogenetic\_Skeleton\_Abundance\_Heatmap.ipynb

### **Conflict of Interest**

The authors declare no conflict of interest.

### **Acknowledgements**

DM, and BH acknowledge support for this project from the NSF Dimensions of Biodiversity Grant (DEB 1737898) and the NSF-funded doctoral student training grant Integrated training Model in Plant And Compu-Tational Sciences (IMPACTS). BH acknowledges funding from the

Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409 and support from the Department of Biochemistry and Molecular Biology startup funding and support from AgBioResearch (BH=MICL02454; GB= MICL02416). BH gratefully acknowledges support from the MSU James K. Billman, Jr. MD endowment. OE is funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC 2048/1, Project ID: 390686111. MvA is funded by EU's Horizon 2020 research and innovation program under the Grant Agreement 862087. We collectively acknowledge that Michigan State University occupies the ancestral, traditional, and contemporary Lands of the Anishinaabeg – Three Fires Confederacy of Ojibwe, Odawa, and Potawatomi peoples. In particular, the University resides on Land ceded in the 1819 Treaty of Saginaw. We recognize, support, and advocate for the sovereignty of Michigan's twelve federally-recognized Indian nations, for historic Indigenous communities in Michigan, for Indigenous individuals and communities who live here now, and for those who were forcibly removed from their Homelands. By offering this Land Acknowledgement, we affirm Indigenous sovereignty and will work to hold Michigan State University more accountable to the needs of American Indian and Indigenous peoples.

## REFERENCES

- Alicandri, E., Paolacci, A. R., Osadolor, S., Sorgonà, A., Badiani, M., & Ciaffi, M. (2020). On the Evolution and Functional Diversity of Terpene Synthases in the Pinus Species: A Review. *Journal of Molecular Evolution*, 88(3), 253–283. <https://doi.org/10.1007/s00239-020-09930-8>
- Appendino, G., Lusso, P., Gariboldi, P., Bombardelli, E., & Gabetta, B. (1992). A 3,11-cyclotaxane from *Taxus baccata*. *Phytochemistry*, 31(12), 4259–4262. [https://doi.org/10.1016/0031-9422\(92\)80455-N](https://doi.org/10.1016/0031-9422(92)80455-N)
- Aros, D., Gonzalez, V., Allemann, R. K., Müller, C. T., Rosati, C., & Rogers, H. J. (2012). Volatile emissions of scented Alstroemeria genotypes are dominated by terpenes, and a myrcene synthase gene is highly expressed in scented Alstroemeria flowers. *Journal of Experimental Botany*, 63(7), 2739–2752. <https://doi.org/10.1093/jxb/err456>
- Arteca, G. A. (1996). Molecular Shape Descriptors. In K. B. Lipkowitz & D. B. Boyd (Eds.), *Reviews in Computational Chemistry* (1st ed., Vol. 9, pp. 191–253). Wiley. <https://doi.org/10.1002/9780470125861.ch5>
- Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., & Werck-Reichhart, D. (2011). Cytochromes P450. *The Arabidopsis Book / American Society of Plant Biologists*, 9, e0144. <https://doi.org/10.1199/tab.0144>
- Bian, G., Han, Y., Hou, A., Yuan, Y., Liu, X., Deng, Z., & Liu, T. (2017). Releasing the potential power of terpene synthases by a robust precursor supply platform. *Metabolic Engineering*, 42, 1–8. <https://doi.org/10.1016/j.ymben.2017.04.006>
- Boncan, D. A. T., Tsang, S. S. K., Li, C., Lee, I. H. T., Lam, H.-M., Chan, T.-F., & Hui, J. H. L. (2020). Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *International Journal of Molecular Sciences*, 21(19), Article 19. <https://doi.org/10.3390/ijms21197382>
- Bryson, A. E., Lanier, E. R., Lau, K. H., Hamilton, J. P., Vaillancourt, B., Mathieu, D., Yocca, A. E., Miller, G. P., Edger, P. P., Buell, C. R., & Hamberger, B. (2023). Uncovering a multiradiene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory. *Nature Communications*, 14, 343. <https://doi.org/10.1038/s41467-023-35845-1>
- Buckingham, J. (2015). *Natural Products Desk Reference*. CRC Press.
- Buckingham, J. (Ed.). (2023). *Dictionary of Natural Products, Supplement 2*. Routledge. <https://doi.org/10.1201/9781315141169>
- Caissard, J.-C., Meekijironroj, A., Baudino, S., & Anstett, M.-C. (2004). Localization of production and emission of pollinator attractant on whole leaves of *Chamaerops humilis*

- (Arecaceae). *American Journal of Botany*, 91(8), 1190–1199.  
<https://doi.org/10.3732/ajb.91.8.1190>
- Carbone, M. R. (2022). When not to use machine learning: A perspective on potential and limitations. *MRS Bulletin*, 47(9), 968–974. <https://doi.org/10.1557/s43577-022-00417-z>
- Celedon, J. M., & Bohlmann, J. (2016). Genomics-Based Discovery of Plant Genes for Synthetic Biology of Terpenoid Fragrances: A Case Study in Sandalwood oil Biosynthesis. *Methods in Enzymology*, 576, 47–67. <https://doi.org/10.1016/bs.mie.2016.03.008>
- Chae, L., Kim, T., Nilo-Poyanco, R., & Rhee, S. Y. (2014). Genomic Signatures of Specialized Metabolism in Plants. *Science*, 344(6183), 510–513.  
<https://doi.org/10.1126/science.1252076>
- Chen, H., Li, G., Köllner, T. G., Jia, Q., Gershenzon, J., & Chen, F. (2014). Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus *Oryza*. *BMC Plant Biology*, 14(1), 239. <https://doi.org/10.1186/s12870-014-0239-x>
- Cheng, C.-H., Lin, Y.-S., Wen, Z.-H., & Su, J.-H. (2012). A new cubitane diterpenoid from the soft coral *Sinularia crassa*. *Molecules (Basel, Switzerland)*, 17(9), 10072–10078.  
<https://doi.org/10.3390/molecules170910072>
- Chou, M.-Y., Andersen, T. B., Mehan Llontop, M. E., Beculheimer, N., Sow, A., Moreno, N., Shade, A., Hamberger, B., & Bonito, G. (2023). Terpenes modulate bacterial and fungal growth and sorghum rhizobiome communities. *Microbiology Spectrum*, 11(5), e01332-23.  
<https://doi.org/10.1128/spectrum.01332-23>
- Croteau, R., Ketchum, R. E. B., Long, R. M., Kaspera, R., & Wildung, M. R. (2006). Taxol biosynthesis and molecular genetics. *Phytochemistry Reviews : Proceedings of the Phytochemical Society of Europe*, 5(1), 75–97. <https://doi.org/10.1007/s11101-005-3748-2>
- Cseke, L. J., Kaufman, P. B., & Kirakosyan, A. (2007). The Biology of Essential Oils in the Pollination of Flowers. *Natural Product Communications*, 2(12), 1934578X0700201225.  
<https://doi.org/10.1177/1934578X0700201225>
- Degenhardt, J., Gershenzon, J., Baldwin, I. T., & Kessler, A. (2003). Attracting friends to feast on foes: Engineering terpene emission to make crop plants more attractive to herbivore enemies. *Current Opinion in Biotechnology*, 14(2), 169–176.  
[https://doi.org/10.1016/S0958-1669\(03\)00025-9](https://doi.org/10.1016/S0958-1669(03)00025-9)
- Degenhardt, J., Köllner, T. G., & Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, 70(15–16), 1621–1637. <https://doi.org/10.1016/j.phytochem.2009.07.030>

- Dötterl, S., & Gershenzon, J. (2023). Chemistry, biosynthesis and biology of floral volatiles: Roles in pollination and other functions. *Natural Product Reports*, 40(12), 1901–1937. <https://doi.org/10.1039/D3NP00024A>
- Durairaj, J., Di Girolamo, A., Bouwmeester, H. J., de Ridder, D., Beekwilder, J., & van Dijk, A. D. (2019). An analysis of characterized plant sesquiterpene synthases. *Phytochemistry*, 158, 157–165. <https://doi.org/10.1016/j.phytochem.2018.10.020>
- Durairaj, J., Melillo, E., Bouwmeester, H. J., Beekwilder, J., Ridder, D. de, & Dijk, A. D. J. van. (2021). Integrating structure-based machine learning and co-evolution to investigate specificity in plant sesquiterpene synthases. *PLOS Computational Biology*, 17(3), e1008197. <https://doi.org/10.1371/journal.pcbi.1008197>
- Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants: A Study in Coevolution. *Evolution*, 18(4), 586–608. <https://doi.org/10.2307/2406212>
- Erbilgin, N., Krokene, P., Christiansen, E., Zeneli, G., & Gershenzon, J. (2006). Exogenous application of methyl jasmonate elicits defenses in Norway spruce (*Picea abies*) and reduces host colonization by the bark beetle *Ips typographus*. *Oecologia*, 148(3), 426–436. <https://doi.org/10.1007/s00442-006-0394-3>
- Fang, C., Fernie, A. R., & Luo, J. (2019). Exploring the Diversity of Plant Metabolism. *Trends in Plant Science*, 24(1), 83–98. <https://doi.org/10.1016/j.tplants.2018.09.006>
- Fay, N., Bliciek, R., Kouklovsky, C., & de la Torre, A. (2022). Total synthesis of grayanane natural products. *Beilstein Journal of Organic Chemistry*, 18, 1707–1719. <https://doi.org/10.3762/bjoc.18.181>
- Feng, Y., Ren, F., Niu, S., Wang, L., Li, L., Liu, X., & Che, Y. (2014). Guanacastane Diterpenoids from the Plant Endophytic Fungus *Cercospora* sp. *Journal of Natural Products*, 77(4 pp.873–881), 873–881.
- Fernández-Martín, R., Reyes, F., Domenech, C. E., Cabrera, E., Bramley, P. M., Barrero, A. F., Avalos, J., & Cerdá-Olmedo, E. (1995). Gibberellin Biosynthesis in *gib* Mutants of *Gibberella fujikuroi*(\*). *Journal of Biological Chemistry*, 270(25), 14970–14974. <https://doi.org/10.1074/jbc.270.25.14970>
- Ferretti, M., Reades, J., & Millington, J. (2019). *Code Camp: 2019 (v1.0)*. <https://doi.org/10.5281/ZENODO.3474043>
- Gao, C., Wang, D., Zhang, Y., Huang, X.-X., & Song, S.-J. (2016). Kaurane and abietane diterpenoids from the roots of *Tripterygium wilfordii* and their cytotoxic evaluation. *Bioorganic & Medicinal Chemistry Letters*, 26(12), 2942–2946. <https://doi.org/10.1016/j.bmcl.2016.04.026>

- Gershenson, J., & Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nature Chemical Biology*, 3(7), Article 7. <https://doi.org/10.1038/nchembio.2007.5>
- González-Coloma, A., Guadaño, A., Tonn, C. E., & Sosa, M. E. (2005). Antifeedant/Insecticidal Terpenes from Asteraceae and Labiatae Species Native to Argentinean Semi-arid Lands. *Zeitschrift Für Naturforschung C*, 60(11–12), 855–861. <https://doi.org/10.1515/znc-2005-11-1207>
- Guenard, D., Gueritte-Voegelein, F., & Potier, P. (1993). Taxol and taxotere: Discovery, chemistry, and structure-activity relationships. *Accounts of Chemical Research*, 26(4), 160–167. <https://doi.org/10.1021/ar00028a005>
- Hao, D.-C. (2021). Drug metabolism and pharmacokinetic diversity of *Taxus* medicinal compounds. In D.-C. Hao (Ed.), *Taxaceae and Cephalotaxaceae* (pp. 123–189). Academic Press. <https://doi.org/10.1016/B978-0-12-823975-9.00004-3>
- Harms, V., Kirschning, A., & S. Dickschat, J. (2020). Nature-driven approaches to non-natural terpene analogues. *Natural Product Reports*, 37(8), 1080–1097. <https://doi.org/10.1039/C9NP00055K>
- Hausch, B. J., Lorjaroenphon, Y., & Cadwallader, K. R. (2015). Flavor chemistry of lemon-lime carbonated beverages. *Journal of Agricultural and Food Chemistry*, 63(1), 112–119. <https://doi.org/10.1021/jf504852z>
- He, D.-H., Matsunami, K., Otsuka, H., Shinzato, T., Aramoto, M., Bando, M., & Takeda, Y. (2005). Tricalysiosides H–O: *Ent*-kaurane glucosides from the leaves of *Tricalysia dubia*. *Phytochemistry*, 66(24), 2857–2864. <https://doi.org/10.1016/j.phytochem.2005.09.014>
- Heiling, S., Schuman, M. C., Schoettner, M., Mukerjee, P., Berger, B., Schneider, B., Jassbi, A. R., & Baldwin, I. T. (2010). Jasmonate and ppHsystemin Regulate Key Malonylation Steps in the Biosynthesis of 17-Hydroxygeranyllinalool Diterpene Glycosides, an Abundant and Effective Direct Defense against Herbivores in *Nicotiana attenuata*. *The Plant Cell*, 22(1), 273–292. <https://doi.org/10.1105/tpc.109.071449>
- Helfrich, E. J. N., Lin, G.-M., Voigt, C. A., & Clardy, J. (2019). Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. *Beilstein Journal of Organic Chemistry*, 15(1), 2889–2906. <https://doi.org/10.3762/bjoc.15.283>
- Hong, Y. J., & Tantillo, D. J. (2011). The Taxadiene-Forming Carbocation Cascade. *Journal of the American Chemical Society*, 133(45), 18249–18256. <https://doi.org/10.1021/ja2055929>
- Hosseini, M., & Pereira, D. M. (2023). The Chemical Space of Terpenes: Insights from Data Science and AI. *Pharmaceuticals*, 16(2), Article 2. <https://doi.org/10.3390/ph16020202>

- Huang, A. C., & Osbourn, A. (2019). Plant terpenes that mediate below-ground interactions: Prospects for bioengineering terpenoids for plant protection. *Pest Management Science*, 75(9), 2368–2377. <https://doi.org/10.1002/ps.5410>
- Jassbi, A. R., Gase, K., Hettenhausen, C., Schmidt, A., & Baldwin, I. T. (2008). Silencing Geranylgeranyl Diphosphate Synthase in *Nicotiana attenuata* Dramatically Impairs Resistance to Tobacco Hornworm. *Plant Physiology*, 146(3), 974–986. <https://doi.org/10.1104/pp.107.108811>
- Jia, Q., Brown, R., Köllner, T. G., Fu, J., Chen, X., Wong, G. K.-S., Gershenzon, J., Peters, R. J., & Chen, F. (2022). Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences*, 119(15), e2100361119. <https://doi.org/10.1073/pnas.2100361119>
- Johnson, S. R., Bhat, W. W., Bibik, J., Turmo, A., Hamberger, B., Evolutionary Mint Genomics Consortium, & Hamberger, B. (2019a). A database-driven approach identifies additional diterpene synthase activities in the mint family (Lamiaceae). *Journal of Biological Chemistry*, 294(4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>
- Johnson, S. R., Bhat, W. W., Sadre, R., Miller, G. P., Garcia, A. S., & Hamberger, B. (2019b). Promiscuous terpene synthases from *Prunella vulgaris* highlight the importance of substrate and compartment switching in terpene synthase evolution. *The New Phytologist*, 223(1), 323–335. <https://doi.org/10.1111/nph.15778>
- J. Tantillo, D. (2011). Biosynthesis via carbocations: Theoretical studies on terpene formation. *Natural Product Reports*, 28(6), 1035–1053. <https://doi.org/10.1039/C1NP00006C>
- Karunanithi, P. S., & Zerbe, P. (2019). Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/articles/10.3389/fpls.2019.01166>
- Keeling, C. I., & Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry*, 67(22), 2415–2423. <https://doi.org/10.1016/j.phytochem.2006.08.019>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter development team. (2016). *Jupyter Notebooks – a publishing format for reproducible computational workflows* (F. Loizides & B. Schmidt, Eds.; pp. 87–90). IOS Press. <https://doi.org/10.3233/978-1-61499-649-1-87>
- Koul, O. (2008). Phytochemicals and Insect Control: An Antifeedant Approach. *Critical Reviews in Plant Sciences*, 27(1), 1–24. <https://doi.org/10.1080/07352680802053908>
- Kumar, J., Gomez-Cano†, F., Hunt†, S. W., Lotreck†, S. G., Mathieu†, D. T., Wilson†, M. L., Long\*, T. M., Kumar\*, J., Gomez-Cano†, F., Hunt†, S. W., Lotreck†, S. G., Mathieu†, D. T., Wilson†, M. L., Long\*, T. M., Kumar\*, J., Gomez-Cano†, F., Hunt†, S. W., Lotreck†, S.

- G., Mathieu†, D. T., ... Long\*, T. M. (2023). *Central Dogma, Dictionaries, and Functions: Using Programming Concepts to Simulate Biological Processes*.  
<https://qubeshub.org/community/groups/coursesource/publications?id=4356>
- Kumar, M. (2021, February 13). *SMILES strings explained for beginners (Cheminformatics Part 1)*. ChemicBook. <https://chemicbook.com/2021/02/13/smiles-strings-explained-for-beginners-part-1.html>
- Kutyna, D. R., & Borneman, A. R. (2018). Heterologous Production of Flavour and Aroma Compounds in *Saccharomyces cerevisiae*. *Genes*, 9(7), Article 7.  
<https://doi.org/10.3390/genes9070326>
- Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R., & Commandeur, U. (2014). Challenges and advances in the heterologous expression of cellulolytic enzymes: A review. *Biotechnology for Biofuels*, 7(1), 135. <https://doi.org/10.1186/s13068-014-0135-5>
- Landrum, G. (2023). *Getting Started with the RDKit in Python*.  
<https://www.rdkit.org/docs/GettingStartedInPython.html>
- Lange, B. M., Mahmoud, S. S., Wildung, M. R., Turner, G. W., Davis, E. M., Lange, I., Baker, R. C., Boydston, R. A., & Croteau, R. B. (2011). Improving peppermint essential oil yield and composition by metabolic engineering. *Proceedings of the National Academy of Sciences*, 108(41), 16944–16949. <https://doi.org/10.1073/pnas.1111558108>
- Laurent, P., Braekman, J.-C., Daloz, D., & Pasteels, J. (2003). Biosynthesis of Defensive Compounds from Beetles and Ants. *European Journal of Organic Chemistry*, 2003(15), 2733–2743. <https://doi.org/10.1002/ejoc.200300008>
- Li, D., Xu, S., Cai, H., Pei, L., Zhang, H., Wang, L., Yao, H., Wu, X., Jiang, J., Sun, Y., & Xu, J. (2013). Enmein-type diterpenoid analogs from natural kaurene-type oridonin: Synthesis and their antitumor biological evaluation. *European Journal of Medicinal Chemistry*, 64, 215–221. <https://doi.org/10.1016/j.ejmech.2013.04.012>
- Li, H., & S. Dickschat, J. (2022). Isotopic labelling experiments and enzymatic preparation of iso-casbenes with casbene synthase from *Ricinus communis*. *Organic Chemistry Frontiers*, 9(3), 795–801. <https://doi.org/10.1039/D1QO01707A>
- Li, J., Halitschke, R., Li, D., Paetz, C., Su, H., Heiling, S., Xu, S., & Baldwin, I. T. (2021). Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science*, 371(6526), 255–260. <https://doi.org/10.1126/science.abe4713>
- Li, Z., & Rudolf, J. D. (2023). Biosynthesis, enzymology, and future of eunicellane diterpenoids. *Journal of Industrial Microbiology & Biotechnology*, 50(1), kuad027.  
<https://doi.org/10.1093/jimb/kuad027>

- Li, Z., Xu, B., Kojasoy, V., Ortega, T., Adressa, D. A., Ning, W., Wei, X., Liu, J., Tantillo, D. J., Loesgen, S., & Rudolf, J. D. (2023). First *trans*-eunicellane terpene synthase in bacteria. *Chem*, 9(3), 698–708. <https://doi.org/10.1016/j.chempr.2022.12.006>
- Liénard, D., & Nogué, F. (2009). *Physcomitrella patens*: A non-vascular plant for recombinant protein production. *Methods in Molecular Biology (Clifton, N.J.)*, 483, 135–144. [https://doi.org/10.1007/978-1-59745-407-0\\_8](https://doi.org/10.1007/978-1-59745-407-0_8)
- Lipińska, M. M., Gołębiowski, M., Szlachetko, D. L., & Kowalkowska, A. K. (2022). Floral attractants in the black orchid *Brasiliorchis schunkeana* (Orchidaceae, Maxillariinae): Clues for presumed sapromyophily and potential antimicrobial activity. *BMC Plant Biology*, 22(1), 575. <https://doi.org/10.1186/s12870-022-03944-8>
- Lu, X., Zhang, J., Brown, B., Li, R., Rodríguez-Romero, J., Berasategui, A., Liu, B., Xu, M., Luo, D., Pan, Z., Baerson, S. R., Gershenzon, J., Li, Z., Sesma, A., Yang, B., & Peters, R. J. (2018). Inferring Roles in Defense from Metabolic Allocation of Rice Diterpenoids. *The Plant Cell*, 30(5), 1119–1131. <https://doi.org/10.1105/tpc.18.00205>
- Luo, D., Callari, R., Hamberger, B., Wubshet, S. G., Nielsen, M. T., Andersen-Ranberg, J., Hallström, B. M., Cozzi, F., Heider, H., Lindberg Møller, B., Staerk, D., & Hamberger, B. (2016). Oxidation and cyclization of casbene in the biosynthesis of Euphorbia factors from mature seeds of *Euphorbia lathyris* L. *Proceedings of the National Academy of Sciences of the United States of America*, 113(34), E5082-5089. <https://doi.org/10.1073/pnas.1607504113>
- Luo, S.-H., Hua, J., Li, C.-H., Jing, S.-X., Liu, Y., Li, X.-N., Zhao, X., & Li, S.-H. (2012). New Antifeedant C20 Terpenoids from *Leucosceptrum canum*. *Organic Letters*, 14(22), 5768–5771. <https://doi.org/10.1021/ol302787c>
- Malik, M. M. (2020). *A Hierarchy of Limitations in Machine Learning* (arXiv:2002.05193). arXiv. <http://arxiv.org/abs/2002.05193>
- McGuire, W. P., Hoskins, W. J., Brady, M. F., Kucera, P. R., Partridge, E. E., Look, K. Y., Clarke-Pearson, D. L., & Davidson, M. (1997). Comparison of combination therapy with paclitaxel and cisplatin versus cyclophosphamide and cisplatin in patients with suboptimal stage III and stage IV ovarian cancer: A Gynecologic Oncology Group study. *Seminars in Oncology*, 24(1 Suppl 2), S2-13-S2-16.
- Meguro, A., Tomita, T., Nishiyama, M., & Kuzuyama, T. (2013). Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton. *Chembiochem: A European Journal of Chemical Biology*, 14(3), 316–321. <https://doi.org/10.1002/cbic.201200651>
- Mendes, E., Ramallete, C., & Duarte, N. (2023). Myrsinane-Type Diterpenes: A Comprehensive Review on Structural Diversity, Chemistry and Biological Activities. *International Journal of Molecular Sciences*, 25(1), 147. <https://doi.org/10.3390/ijms25010147>

- Miller, G. P., Bhat, W. W., Lanier, E. R., Johnson, S. R., Mathieu, D. T., & Hamberger, B. (2020). The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all-*cis* prenyl intermediate. *The Plant Journal*, *104*(3), 693–705. <https://doi.org/10.1111/tpj.14957>
- Miyazaki, S., Nakajima, M., & Kawaide, H. (2015). Hormonal diterpenoids derived from *ent*-kaurenoic acid are involved in the blue-light avoidance response of *Physcomitrella patens*. *Plant Signaling & Behavior*, *10*(2), e989046. <https://doi.org/10.4161/15592324.2014.989046>
- Moon, N. G., & Harned, A. M. (2018). Synthetic explorations of the briarane jungle: Progress in developing a synthetic route to a common family of diterpenoid natural products. *Royal Society Open Science*, *5*(5), 172280. <https://doi.org/10.1098/rsos.172280>
- Nagel, R., Berasategui, A., Paetz, C., Gershenzon, J., & Schmidt, A. (2014). Overexpression of an Isoprenyl Diphosphate Synthase in Spruce Leads to Unexpected Terpene Diversion Products That Function in Plant Defense. *Plant Physiology*, *164*(2), 555–569. <https://doi.org/10.1104/pp.113.228940>
- Ndi, C. P., Semple, S. J., Griesser, H. J., Pyke, S. M., & Barton, M. D. (2007). Antimicrobial compounds from the Australian desert plant *Eremophila neglecta*. *Journal of Natural Products*, *70*(9), 1439–1443. <https://doi.org/10.1021/np070180r>
- Nuutinen, T. (2018). Medicinal properties of terpenes found in *Cannabis sativa* and *Humulus lupulus*. *European Journal of Medicinal Chemistry*, *157*, 198–228. <https://doi.org/10.1016/j.ejmech.2018.07.076>
- O’Boyle, N. M. (2012). Towards a Universal SMILES representation—A standard method to generate canonical SMILES based on the InChI. *Journal of Cheminformatics*, *4*(1), 22. <https://doi.org/10.1186/1758-2946-4-22>
- O’Donnell, T. J., Rao, S. N., Koehler, K., Martin, Y. C., & Eccles, B. (1991). A general approach for atom-type assignment and the interconversion of molecular structure files. *Journal of Computational Chemistry*, *12*(2), 209–214. <https://doi.org/10.1002/jcc.540120210>
- Pan, S., Chen, S., & Dong, G. (2018). Divergent Total Syntheses of Enmein-Type Natural Products: (–)-Enmein, (–)-Isodocarpin, and (–)-Sculponin R. *Angewandte Chemie International Edition*, *57*(21), 6333–6336. <https://doi.org/10.1002/anie.201803709>
- Pateraki, I., Andersen-Ranberg, J., Jensen, N. B., Wubshet, S. G., Heskes, A. M., Forman, V., Hallström, B., Hamberger, B., Motawia, M. S., Olsen, C. E., Staerk, D., Hansen, J., Møller, B. L., & Hamberger, B. (2017). Total biosynthesis of the cyclic AMP booster forskolin from *Coleus forskohlii*. *eLife*, *6*, e23001. <https://doi.org/10.7554/eLife.23001>
- Peters, R. J. (2010). Two rings in them all: The labdane-related diterpenoids. *Natural Product Reports*, *27*(11), 1521. <https://doi.org/10.1039/c0np00019a>

- Philippe, R. N., De Mey, M., Anderson, J., & Ajikumar, P. K. (2014). Biotechnological production of natural zero-calorie sweeteners. *Current Opinion in Biotechnology*, *26*, 155–161. <https://doi.org/10.1016/j.copbio.2014.01.004>
- Piccoli, P., & Bottini, R. (2013). Terpene Production by Bacteria and its Involvement in Plant Growth Promotion, Stress Alleviation, and Yield Increase. In *Molecular Microbial Ecology of the Rhizosphere* (pp. 335–343). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118297674.ch31>
- Pichersky, E., & Lewinsohn, E. (2011). Convergent Evolution in Plant Specialized Metabolism. *Annual Review of Plant Biology*, *62*(1), 549–566. <https://doi.org/10.1146/annurev-arplant-042110-103814>
- Proffit, M., Lapeyre, B., Buatois, B., Deng, X., Arnal, P., Gouzerh, F., Carrasco, D., & Hossaert-McKey, M. (2020). Chemical signal is in the blend: Bases of plant-pollinator encounter in a highly specialized interaction. *Scientific Reports*, *10*(1), Article 1. <https://doi.org/10.1038/s41598-020-66655-w>
- Reddy, G. K., Leferink, N. G. H., Umemura, M., Ahmed, S. T., Breitling, R., Scrutton, N. S., & Takano, E. (2020). Exploring novel bacterial terpene synthases. *PLOS ONE*, *15*(4), e0232220. <https://doi.org/10.1371/journal.pone.0232220>
- Rinkel, J., Lauterbach, L., Rabe, P., & Dickschat, J. S. (2018). Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria albata*. *Angewandte Chemie International Edition*, *57*(12), 3238–3241. <https://doi.org/10.1002/anie.201800385>
- Rinkel, J., Rabe, P., Chen, X., Köllner, T. G., Chen, F., & Dickschat, J. S. (2017). Mechanisms of the Diterpene Cyclases  $\beta$ -Pinacene Synthase from *Dictyostelium discoideum* and Hydropyrene Synthase from *Streptomyces clavuligerus*. *Chemistry (Weinheim an Der Bergstrasse, Germany)*, *23*(44), 10501–10505. <https://doi.org/10.1002/chem.201702704>
- Schalk, M., Pastore, L., Mirata, M. A., Khim, S., Schouwey, M., Deguerry, F., Pineda, V., Rocci, L., & Daviet, L. (2012). Toward a Biosynthetic Route to Sclareol and Amber Odorants. *Journal of the American Chemical Society*, *134*(46), 18900–18903. <https://doi.org/10.1021/ja307404u>
- Schiebe, C., Hammerbacher, A., Birgersson, G., Witzell, J., Brodelius, P. E., Gershenzon, J., Hansson, B. S., Krokene, P., & Schlyter, F. (2012). Inducibility of chemical defenses in Norway spruce bark is correlated with unsuccessful mass attacks by the spruce bark beetle. *Oecologia*, *170*(1), 183–198. <https://doi.org/10.1007/s00442-012-2298-8>
- Schneider, F., Pan, L., Ottenbruch, M., List, T., & Gaich, T. (2021). The Chemistry of Nonclassical Taxane Diterpene. *Accounts of Chemical Research*, *54*(10), 2347–2360. <https://doi.org/10.1021/acs.accounts.0c00873>

- Schrepfer, P., Ugur, I., Klumpe, S., Loll, B., Kaila, V. R. I., & Brück, T. (2020). Exploring the catalytic cascade of cembranoid biosynthesis by combination of genetic engineering and molecular simulations. *Computational and Structural Biotechnology Journal*, 18, 1819–1829. <https://doi.org/10.1016/j.csbj.2020.06.030>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Shebek, K. M., Strutz, J., Broadbelt, L. J., & Tyo, K. E. J. (2023). Pickaxe: A Python library for the prediction of novel metabolic reactions. *BMC Bioinformatics*, 24(1), 106. <https://doi.org/10.1186/s12859-023-05149-8>
- Strutz, J., Shebek, K. M., Broadbelt, L. J., & Tyo, K. E. J. (2022). MINE 2.0: Enhanced biochemical coverage for peak identification in untargeted metabolomics. *Bioinformatics*, 38(13), 3484–3487. <https://doi.org/10.1093/bioinformatics/btac331>
- Tantillo, D. J. (2010). The carbocation continuum in terpene biosynthesis—Where are the secondary cations? *Chemical Society Reviews*, 39(8), 2847–2854. <https://doi.org/10.1039/B917107J>
- Tantillo, D. J. (2011). Biosynthesis via carbocations: Theoretical studies on terpene formation. *Natural Product Reports*, 28(6), 1035–1053. <https://doi.org/10.1039/C1NP00006C>
- Tejaro, C. N., Adhikari, A., & Shen, B. (2019). Challenges and opportunities for natural product discovery, production, and engineering in native producers versus heterologous hosts. *Journal of Industrial Microbiology and Biotechnology*, 46(3–4), 433–444. <https://doi.org/10.1007/s10295-018-2094-5>
- Tetali, S. D. (2019). Terpenes and isoprenoids: A wealth of compounds for global use. *Planta*, 249(1), 1–8. <https://doi.org/10.1007/s00425-018-3056-x>
- Theis, N., & Lerda, M. (2003). The Evolution of Function in Plant Secondary Metabolites. *International Journal of Plant Sciences*, 164(S3), S93–S102. <https://doi.org/10.1086/374190>
- Tice, B. S. (2014). Line Notation Systems and Compression. In *Algorithmic Techniques for the Polymer Sciences*. Apple Academic Press.
- Todeschini, R., & Consonni, V. (2003). Descriptors from Molecular Geometry. In J. Gasteiger (Ed.), *Handbook of Chemoinformatics* (1st ed., pp. 1004–1033). Wiley. <https://doi.org/10.1002/9783527618279.ch37>
- Tong, Y., Hu, T., Tu, L., Chen, K., Liu, T., Su, P., Song, Y., Liu, Y., Huang, L., & Gao, W. (2021). Functional characterization and substrate promiscuity of sesquiterpene synthases

from *Tripterygium wilfordii*. *International Journal of Biological Macromolecules*, 185, 949–958. <https://doi.org/10.1016/j.ijbiomac.2021.07.004>

Toropov, A. A., Toropova, A. P., Mukhamedzhanova, D. V., & Gutman, I. (2005). Simplified molecular input line entry system (SMILES) as an alternative for constructing quantitative structure-property relationships (QSPR). *IJC-A Vol.44A(08) [August 2005]*. <http://nopr.niscpr.res.in/handle/123456789/18068>

Toyomasu, T., Usui, M., Sugawara, C., Otomo, K., Hirose, Y., Miyao, A., Hirochika, H., Okada, K., Shimizu, T., Koga, J., Hasegawa, M., Chuba, M., Kawana, Y., Kuroda, M., Minami, E., Mitsuhashi, W., & Yamane, H. (2014). Reverse-genetic approach to verify physiological roles of rice phytoalexins: Characterization of a knockdown mutant of OsCPS4 phytoalexin biosynthetic gene in rice. *Physiologia Plantarum*, 150(1), 55–62. <https://doi.org/10.1111/ppl.12066>

Turlik, A., Chen, Y., Scruse, A. C., & Newhouse, T. R. (2019). Convergent Total Synthesis of Principinol D, a Rearranged Kaurane Diterpenoid. *Journal of the American Chemical Society*, 141(20), 8088–8092. <https://doi.org/10.1021/jacs.9b03751>

Ujita, E. (n.d.). *Total Synthesis of Enrnein*.

Van Drie, J. H., Weininger, D., & Martin, Y. C. (1989). ALADDIN: An integrated tool for computer-assisted molecular design and pharmacophore recognition from geometric, steric, and substructure searching of three-dimensional molecular structures. *Journal of Computer-Aided Molecular Design*, 3(3), 225–251. <https://doi.org/10.1007/BF01533070>

Villegas-Plazas, M., Wos-Oxley, M. L., Sanchez, J. A., Pieper, D. H., Thomas, O. P., & Junca, H. (2019). Variations in Microbial Diversity and Metabolite Profiles of the Tropical Marine Sponge *Xestospongia muta* with Season and Depth. *Microbial Ecology*, 78(1), 243–256. <https://doi.org/10.1007/s00248-018-1285-y>

Wang, G., Tang, W., & Bidigare, R. R. (2005). Terpenoids As Therapeutic Drugs and Pharmaceutical Agents. In L. Zhang & A. L. Demain (Eds.), *Natural Products: Drug Discovery and Therapeutic Medicine* (pp. 197–227). Humana Press. [https://doi.org/10.1007/978-1-59259-976-9\\_9](https://doi.org/10.1007/978-1-59259-976-9_9)

Wang, Z., Nelson, D., Zhang, J., Wan, X., & J. Peters, R. (2023). Plant (di)terpenoid evolution: From pigments to hormones and beyond. *Natural Product Reports*, 40(2), 452–469. <https://doi.org/10.1039/D2NP00054G>

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93(9), 2325–2327. <https://doi.org/10.1021/ja00738a045>

- Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1), 31–36. <https://doi.org/10.1021/ci00057a005>
- Weininger, D. (1990). SMILES. 3. DEPICT. Graphical depiction of chemical structures. *Journal of Chemical Information and Computer Sciences*, 30(3), 237–243. <https://doi.org/10.1021/ci00067a005>
- Weininger, D., Weininger, A., & Weininger, J. L. (1989). SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences*, 29(2), 97–101. <https://doi.org/10.1021/ci00062a008>
- Weng, J.-K. (2014). The evolutionary paths towards complexity: A metabolic perspective. *New Phytologist*, 201(4), 1141–1149. <https://doi.org/10.1111/nph.12416>
- 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(10), 1037–1045. <https://doi.org/10.1038/s41589-021-00822-6>
- Wilson, S. A., & Roberts, S. C. (2012). Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant Biotechnology Journal*, 10(3), 249–268. <https://doi.org/10.1111/j.1467-7652.2011.00664.x>
- Wong, J., de Rond, T., d'Espaux, L., van der Horst, C., Dev, I., Rios-Solis, L., Kirby, J., Scheller, H., & Keasling, J. (2018). High-titer production of lathyrane diterpenoids from sugar by engineered *Saccharomyces cerevisiae*. *Metabolic Engineering*, 45, 142–148. <https://doi.org/10.1016/j.ymben.2017.12.007>
- Xu, B., Ning, W., Wei, X., & D. Rudolf, J. (2022). Mutation of the eunicellane synthase Bnd4 alters its product profile and expands its prenylation ability. *Organic & Biomolecular Chemistry*, 20(45), 8833–8837. <https://doi.org/10.1039/D2OB01931K>
- Yan, X.-Y., Zhang, L., Yang, Q.-B., Ge, Z.-Y., Liang, L.-F., & Guo, Y.-W. (2023). Genus *Litophyton*: A Hidden Treasure Trove of Structurally Unique and Diversely Bioactive Secondary Metabolites. *Marine Drugs*, 21(10), Article 10. <https://doi.org/10.3390/md21100523>
- Yang, J., Wang, W.-G., Wu, H.-Y., Du, X., Li, X.-N., Li, Y., Pu, J.-X., & Sun, H.-D. (2016). Bioactive Enmein-Type ent-Kaurane Diterpenoids from *Isodon phyllostachys*. *Journal of Natural Products*, 79(1), 132–140. <https://doi.org/10.1021/acs.jnatprod.5b00802>
- Yoon, H. S., Andersen, R. A., Boo, S. M., & Bhattacharya, D. (2009). Stramenopiles. In M. Schaechter (Ed.), *Encyclopedia of Microbiology (Third Edition)* (pp. 721–731). Academic Press. <https://doi.org/10.1016/B978-012373944-5.00253-4>

- Zahran, E. M., Abdelmohsen, U. R., Ayoub, A. T., Salem, M. A., Khalil, H. E., Desoukey, S. Y., Fouad, M. A., & Kamel, M. S. (2020). Metabolic profiling, histopathological anti-ulcer study, molecular docking and molecular dynamics of ursolic acid isolated from *Ocimum forskolei* Benth. (Family Lamiaceae). *South African Journal of Botany*, *131*, 311–319. <https://doi.org/10.1016/j.sajb.2020.03.004>
- Zeng, T., Chen, Y., Jian, Y., Zhang, F., & Wu, R. (2022). Chemotaxonomic investigation of plant terpenoids with an established database (TeroMOL). *New Phytologist*, *235*(2), 662–673. <https://doi.org/10.1111/nph.18133>
- Zeng, T., Hess, Jr., Bernard Andes, Zhang, F., & Wu, R. (2022). Bio-inspired chemical space exploration of terpenoids. *Briefings in Bioinformatics*, *23*(5), bbac197. <https://doi.org/10.1093/bib/bbac197>
- Zeng, T., Liu, Z., Liu, H., He, W., Tang, X., Xie, L., & Wu, R. (2019). Exploring Chemical and Biological Space of Terpenoids. *Journal of Chemical Information and Modeling*, *59*(9), 3667–3678. <https://doi.org/10.1021/acs.jcim.9b00443>
- Zeng, T., Liu, Z., Zhuang, J., Jiang, Y., He, W., Diao, H., Lv, N., Jian, Y., Liang, D., Qiu, Y., Zhang, R., Zhang, F., Tang, X., & Wu, R. (2020). TeroKit: A Database-Driven Web Server for Terpenome Research. *Journal of Chemical Information and Modeling*, *60*(4), 2082–2090. <https://doi.org/10.1021/acs.jcim.0c00141>
- Zerbe, P., & Bohlmann, J. (2014). Bioproducts, Biofuels, and Perfumes: Conifer Terpene Synthases and their Potential for Metabolic Engineering. In R. Jetter (Ed.), *Phytochemicals – Biosynthesis, Function and Application: Volume 44* (pp. 85–107). Springer International Publishing. [https://doi.org/10.1007/978-3-319-04045-5\\_5](https://doi.org/10.1007/978-3-319-04045-5_5)
- Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R., & Bohlmann, J. (2012). Bifunctional *cis*-Abienol Synthase from *Abies balsamea* Discovered by Transcriptome Sequencing and Its Implications for Diterpenoid Fragrance Production. *The Journal of Biological Chemistry*, *287*(15), 12121–12131. <https://doi.org/10.1074/jbc.M111.317669>
- Zhang, A., Xiong, Y., Fang, J., Jiang, X., Wang, T., Liu, K., Peng, H., & Zhang, X. (2022). Diversity and Functional Evolution of Terpene Synthases in Rosaceae. *Plants*, *11*(6), Article 6. <https://doi.org/10.3390/plants11060736>
- Zhao, D.-D., Jiang, L.-L., Li, H.-Y., Yan, P.-F., & Zhang, Y.-L. (2016). Chemical Components and Pharmacological Activities of Terpene Natural Products from the Genus *Paeonia*. *Molecules (Basel, Switzerland)*, *21*(10), 1362. <https://doi.org/10.3390/molecules21101362>
- Zhao, T., Krokene, P., Hu, J., Christiansen, E., Björklund, N., Långström, B., Solheim, H., & Borg-Karlson, A.-K. (2011). Induced Terpene Accumulation in Norway Spruce Inhibits Bark Beetle Colonization in a Dose-Dependent Manner. *PLOS ONE*, *6*(10), e26649. <https://doi.org/10.1371/journal.pone.0026649>

Zi, J., & Peters, R. J. (2013). Characterization of CYP76AH4 clarifies phenolic diterpenoid biosynthesis in the Lamiaceae. *Organic & Biomolecular Chemistry*, 11(44), 7650–7652. <https://doi.org/10.1039/c3ob41885e>

Zou, J., Ye, J., Zhao, C., Zhang, J., Liu, Y., Pan, L., He, K., & Zhang, H. (2023). Guidongnins I–J: Two New 6,7-seco-7,20-Olide-ent-kaurene Diterpenes with Unusual Structures from *Isodon rubescens*. *International Journal of Molecular Sciences*, 24(17), Article 17. <https://doi.org/10.3390/ijms241713451>

**CHAPTER 4**  
**Long Terminal Repeat Retrotransposon Targeted Transformation and Development of Promoter Reporter System in *Physcomitrium patens* for Sequential Targeting of Diterpene Module**

Mathieu, D., Banerjee, A., Motsa, B., Hamberger, B.

## Abstract

The bryophyte *Physcomitrium patens* shows potential as an effective system for heterologous production of specialized metabolites at an industrial scale. Of these high value metabolites, diterpenoids have been a major draw, due to their application in flavors, fragrances, feedstocks, and pharmaceuticals to name a few. *P. patens* has many advantages as a chassis for the heterologous expression of metabolites because of its ability to form genetically stable transformants, many -omic libraries, and containing protein assembly machinery closely related to vascular plants (the primary source of diterpenes). Transformation is typically targeted to the Pp108 locus within the *P. patens* genome, due to its neutral phenotype when knocked out and the common success rate at which it undergoes transformation, despite transformants often showing low expression. Here we build upon a transformation library generated through a shotgun approach for transformation where the Gypsy long terminal repeat retrotransposons (RLG2) were targeted, providing a wide portfolio of enhanced yellow fluorescent protein (YFP) expression. Unlike most organisms, transposable elements (TEs) in the *P. patens* genome are homogeneously distributed throughout both euchromatic and heterochromatic regions, making this system one of few that may provide effective and reliable expression levels at these sites. Additionally, most synthetic biological tools in plants utilize the ubiquitin promoter originating from *Zea mays*, however regulatory elements are especially prone to changes in motif recognition and mutation. Identifying regulatory elements sourced from *P. patens* would provide opportunity for higher expression overall and greater control for conditional gene expression. Here, a reporter system is developed for easy insertion and testing of promoter elements upstream of an eYFP gene, with homologous regions to the Pp108 loci. Unconventional approaches for improving *P. patens* expression of heterologous genes by targeting alternative loci

or testing alternative promoters provides direction for improving the system as a whole for the production of high-value metabolites in the future.

## **Keywords**

*Physcomitrium patens*, synthetic engineering, homologous recombination, long terminal repeat retrotransposons, reporter systems, promoters, diterpenes

## **Introduction**

Terpenoids make up one of the most structurally diverse and abundant groups of natural products, with over 160,000 known compounds to date [Buckingham 2015, Christianson 2017, Zeng et al. 2020, Zhou & Pichersky 2020, Buckingham 2024]. The diterpene sub-family consists of over 60,000 unique compounds to-date and most are derived from the 20 Carbon precursor geranylgeranyl diphosphate (GGDP) [Heusler 1902, Zerbe et al. 2013, Johnson et al. 2019, Zeng et al. 2020]. The current abundance of chemodiversity likely can be attributed to their role in communication and defense, which face constant selection pressure in nature. The diterpene synthase (diTPS) enzymes are responsible for the early stages of synthesis and because most products influence specialized metabolism allows these enzyme families more flexibility for natural expansion and neofunctionalization [Buckingham 2015]. As the largest producers of diterpenes, land plants find utility from them in organism defense, pollinator attraction, developmental signaling, and interspecies communication [Degenhardt et al. 2003; Theis and Lerdau 2003; Aros et al. 2012; Boncan et al. 2020; Caissard et al. 2004; Chou et al. 2023; Cseke et al. 2007; Gershenzon & Dudareva 2007; Dötterl & Gershenzon 2023; Erbilgin et al., 2006; Jassbi et al. 2008; Heiling et al. 2010; Huang & Osbourn 2019; Keeling and Bohlmann, 2006;

Laurent et al. 2003; Li et al. 2021; Lipińska et al. 2022; Lu et al. 2018; Miller et al. 2020; Miyazaki et al. 2015; Nagel et al. 2014; Ndi et al. 2007; Piccoli & Bottini 2013; Proffit et al. 2020; Schiebe et al., 2012; Toyomasu et al. 2014; Wang et al. 2023; Zhao et al., 2011]. In addition to their role in nature, these compounds have significant value in humanitarian applications as well with their use as fragrances, flavors, fuels, pesticides, nutraceuticals, and pharmaceuticals [Degenhardt et al. 2003; González-Coloma et al. 2014; Hausch et al. 2015; Koul 2008; Lange et al. 2011; Schalk et al. 2011; Phillipe et al. 2014; Celedon & Bohlmann 2016; Kutiyana & Bornemann 2018; Nuutinen 2018; Tetali 2019; Wang et al. 2005; Wani et al. 1971; Wilson & Roberts 2011; Zerbe et al. 2012; Zerbe and Bohlmann 2014; Zhao et al. 2016]. However, cultivating these compounds in vivo often demands significant resources, space, and time, with production often suboptimal in their native host [Lambertz et al. 2014, Teijaro et al. 2019]. Therefore, the heterologous expression of diterpene pathways can provide route for huge improvements to production and yield.

The model moss, *Physcomitrium patens* (*P. patens*) provides an excellent framework for the heterologous expression of diterpenes. As previously mentioned most known diterpenes are synthesized in land plants, largely the mints (Lamiaceae), spurge (Euphorbiaceae), and composite flower (Asteraceae) families [Johnson et al. 2019, Zeng et al. 2020]. Because *P. patens* has cellular machinery, physiology, and compartments more similar to native systems provide it huge advantages over other heterologous systems like *Escherichia coli* (*E. coli*; bacteria), *Saccharomyces cerevisiae* (*S. cerevisiae*; yeast), and *Chlamydomonas reinhardtii* (*C. reinhardtii*; algae) [Fang et al. 2019, Liénard & Nogué 2009]. Additionally, chemical diversity within *P. patens* is comparatively low, with only one endogenous bi-functional Class II/Class I diTPS copalyl diphosphate/kaurene synthase (CPS/KS) enzyme, responsible for synthesizing

*ent*-kaurenoic acid directly from GGDP [Hayashi et al. 2006, Hoffmann et al. 2014; Zhan et al. 2014]. Despite *ent*-kauranic acid reaching extremely high concentrations of about 0.37-fold of chlorophyll, CPS/KS knockouts still generates viable cultures [Hoffmann et al. 2014, Zhan et al. 2014]. Although these knockout lines are developmentally stunted as chloronema (non-leafy undifferentiated, filamentous tissue), preventing sexual maturity, these tissue cultures can still be propagated indefinitely with no known adverse effects to plant health [Hayashi et al. 2010].

These mutant lines are perfect for the synthetic production of nonnative diterpenes because they still produce an excess pool of the GGDP precursor. This characteristic has been used previously for production of 13R-*ent*-manoyl oxide, which is a precursor for the weight loss drug forskolin, and taxadiene, which is a precursor to the anticancer drug Taxol™ [Anterola et al. 2009, Pateraki et al. 2014, Bach et al. 2014, Banerjee et al. 2019]. Benefits from culturing *P. patens* come from flexible propagation possible on agar, liquid, and soil medias, capacity to store mutants for long periods of time with cryopreservation, and potential for commercial production without requiring acres of land [Schulte & Reski 2004, Ikram 2023, Mathieu et al. 2024].

Pairing the capacity for stable transformation via homologous recombination with an updated *P. patens* chromosome-scale genome assembly in 2018 provided new opportunities for improving old technologies [Schaefer & Zrýd 1997, Novikova et al. 2008, Lang et al. 2018, Banerjee et al. 2019]. Historically the Pp108 locus has been a classic target for homologous recombination in *P. patens* due to its neutrality when replaced [Schaefer & Zrýd 1997]. While the Pp108 locus is reliable, the *P. patens* genome assembly illustrates the potential for alternative insertion sites, uniquely including the multiple genomic loci for long terminal repeat retrotransposons (LTR-RT) instead [Lang et al. 2018, Banerjee et al. 2019, Vendrell-Mir et al. 2020]. The unique and homogenous distribution of LTR-RTs and relatively high LTR-RT activity provided a unique

opportunity to use a “shotgun approach” for untargeted transformation throughout the genome [Banerjee et al. 2019]. Previous work, replacing LTR-RT loci with an enhanced yellow fluorescent protein (eYFP) at random, was used to generate a panel of eYFP mutants with a range of fluorescent expressions [Lang et al. 2018, Banerjee et al. 2019, Vendrell-Mir et al. 2020]. This illustrates a high variability of expression regarding just the position within the genome, with some of the mutant lines even exceeding eYFP activity from former Pp108 eYFP mutants [Banerjee et al. 2019]. Work done here aims to perform a secondary knockout, now instead targeting the inserted LTR-RT-eYFP loci, to determine if this range of expression correlates when producing the diterpene manoyl oxide instead [Pateraki et al. 2014, Banerjee et al. 2019]. Additionally, a reporter system was developed for testing native *P. patens* promoters activity in native but alternative locations. This will allow promoter activity to be tested in the context of environmental response and to identify native promoters that outcompete the commonly used *Zea mays* ubiquitin promoter, which diverged from *P. patens* 450 million years ago [Christensen & Quail 1996]. The promoter elements intended for testing here captured the upstream region of genes with distinct expression patterns when in the presence of *Benniella erionia* or *Linnemannia elongata* for determining the regulatory capacity of fungal induced response in *P. patens* [Mathieu et al. 2024]. These reporter constructs with inserted promoters would be targeted for the Pp108 locus to compare expression to a neutral background [Banerjee et al. 2019]. This work, while incomplete, provides an important foundation for developing new synthetic biological tools in *P. patens* through alternative loci transformation targeting and identification of novel regulatory regions within the *P. patens* genome.

## Materials & Methods

### *Physcomitrium patens culturing and propagation*

Propagation of moss was performed similarly to previously published work [Cove 2005, Banerjee et al. 2019]. Moss was cultured in a laminar flow hood under sterile conditions. Tissue was added to 25 mL Erlenmeyer flasks with 5 mL autoclaved water, before tissue was blended until homogenous. The mixture was then dispersed evenly onto BCD-media plates (composition: 45  $\mu$ M iron (II) sulfate heptahydrate ( $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ), 1 mM magnesium sulfate ( $\text{MgSO}_4$ ), 1.84 mM monopotassium phosphate ( $\text{KH}_2\text{PO}_4$ ), 10 mM potassium nitrate ( $\text{KNO}_3$ ), trace element solution (1000 $\times$  dilution), 1 mM calcium chloride ( $\text{CaCl}_2$ ), 5 mM diammonium tartrate ( $(\text{NH}_4)_2\text{C}_4\text{H}_4\text{O}_6$ ), agar (0.7% (w/v). Trace element solution,  $\text{Al}_2(\text{SO}_4)_3 \cdot \text{K}_2\text{SO}_4 \cdot 24\text{H}_2\text{O}$ ,  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ ,  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ,  $\text{H}_3\text{BO}_3$ , KBr, KI, LiCl,  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ ,  $\text{SnCl}_2 \cdot 2\text{H}_2\text{O}$ ,  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ). Calcium chloride and diammonium tartrate were added to the BCD agar media immediately before use. Plates were then sealed with micropore tape (3M, VWR) and were incubated at 18h light: 6h dark cycles under LED lights at 100-150  $\mu\text{mol m}^{-2} \text{s}^{-1}$  intensity.

### *PCR-based Assembly of diterpene construct for targeted insertion to LTR-RT/eYFP*

The module targeted to replace LTR-RT/eYFP mutant lines via homologous recombination consisted of one continuous module that contained homogeneous regions to the neomycin phosphotransferase II (NPT-II) herbicide resistance marker, a sulfonamide resistance cassette, actin promoter, a class II diTPS and class I diTPS linked by LP4-2A linker, and octopine synthase gene terminator (OCS-T; Figure 4.1). This utilized individual parts previously reported in Banerjee et al. 2019. Modules were assembled in two steps. The first module, which contained homogenous NPT-II herbicide resistance marker, sulfonamide resistance cassette, and actin promoter were first assembled with each portion containing 20-bp overhangs to their respective

neighbor and were connected using the In-Fusion protocol. The second module was then stitched together, which consisted of the class II/class I diTPS and OCS-T cassettes, by again designing 20bp overhangs between modules and connecting elements with a second In-Fusion protocol. The entirety of this sequence was transformed into a pEAQ vector for long term storage at -20°C [Sainsbury et al. 2009].

### ***PCR-based Assembly of Empty Reporter Gene Construct***

Fragments for the reporter system were acquired as three modules from work previously performed [Banerjee et al. 2019]. The first module contained the 5' region homogenous to the Pp108 locus, NPT-II herbicide resistance marker, and *Zea mays* ubiquitin promoter (PNZ module). This first module (PNZ module) was modified to include a 20bp pEAQ overhang on the 5' Pp108 region and a constructed Cfr9I restriction enzyme site and eYFP overhang, connecting to the NPT-II marker, with the ubiquitin promoter intentionally removed. These alterations generated the first module. The second module (YFP module) contained the full eYFP sequence but with primers designed with 20bp overhang before the start site, including Cfr9I and NPT-II marker, and a 20bp overhang coinciding with the terminator sequence. The third module (OCS module) contained an OCS terminator and the 3' Pp108 homologous region, with 20 bp overhang aligning to eYFP stop site and pEAQ. The PNZ and YFP modules were fused by PCR forming one PNZ-YFP module. In-Fusion cloning was performed to connect the PNZ-YFP module, OCS module, and pEAQ vector [Sainsbury et al. 2009]. After In-Fusion positive antibiotic clones were screened and confirmed for all three PNZ, YFP, and OCS modules independently. Successful clones were validated with sequencing and stored at -20°C.

### ***Identifying Promoter Regions from P. patens***

Promoters were chosen based on measured RNA-seq responses of *P. patens* cocultured with *L. elongata* and *B. erionia* [Mathieu et al. 2024]. Promoters were chosen based on differentially expressed genes that were statistically significant, had interesting physiological implications, and distinct upregulation in the presence of fungi. In order to ensure this response was binary (on or off in the presence of fungi) the candidate gene promoters were selected only when expression in the presence of fungi was 8TPM or higher and was 0 TPM or was nearly 0 TPM when in the absence of fungi.

Tentative promoters were identified in the *P. patens* v3.3 genome in Phytozome by searching gene candidates and comparing 5kb upstream with and without the 5' UTR [Lang et al. 2018]. These regions were shortened in cases where a restriction site for Cfr9I was present earlier or a gene neighbored within that 5kb window. To differentiate the regulatory effect of the 5' UTR and sequence upstream of the UTR, two versions of the Fasta file were generated, where one included the UTR and the other replaced the full UTR with the character "N" so positionally both sequences aligned in relation to the start site. All tentative promoter sequences were evaluated for transcription factor binding sites in PlantPan (v3.0) [Chow et al. 2019] and PlantRegMap/PlantTFDB (v5.0) [Jin et al. 2017]. Results from PlantRegMap/PlantTFDB provided regulatory information more specific to *P. patens* and thus was prioritized. Regions for cloning were chosen based on capturing the highest predicted regulatory elements in the shortest number of base pairs. These values were also adjusted to accommodate AT rich regions for primer design. All primer sets were selected based on similar melting temperatures, acceptable range of GC-content (35-60%), having only one matching BLAST hit within the *P. patens* genome, and with overhangs to include the restriction site for the Cfr9I endonuclease.

### ***PCR amplification of P. patens Promoter Regions and Reporter Gene Assembly***

High molecular weight genomic DNA was isolated from *P. patens* (Gransden 2004) by flash freezing in liquid N<sub>2</sub> and then grinding tissue with mortar and pestle. Ground tissue was placed in 1.5 mL Eppendorf tubes with 500 µL Cetyltrimethylammonium bromide (CTAB) buffer (composition: 2% CTAB, 1% polyvinylpyrrolidone, 100mM Tris·HCl, 1.4 M NaCl, 20 mM Ethylenediaminetetraacetic acid (EDTA)) with 100 mg of ground tissue and was incubated at 60°C for 30 minutes. Extract was centrifuged for 5 minutes at 14,000 x g and supernatant was transferred to a new 1.5 mL Eppendorf tube. 5 µL RNase A was added and sample was incubated at 37°C for 20 minutes. An equal volume of phenol/chloroform/isoamyl alcohol solution (25:24:1) was added, then sample was vortexed and centrifuged for 1 minute at 14,000 x g. The aqueous top layer was then transferred to a new tube. Extractions with phenol/chloroform/isoamyl alcohol, vortexing, centrifuging, and extraction of aqueous layer was repeated until clear. A ratio of 0.7 volume cold isopropanol (-20°C) was added to the extract, then the tube was inverted, and sat at -20°C for 15 minutes. The sample was centrifuged at 14,000 x g for 10 minutes and decanted, leaving the gDNA pellet. The pellet was washed with 500 µL 70% ethanol, centrifuged at 14,000 x g, decanted, and set out to dry for 30 minutes. Final pellet was dissolved in 100 µL Tris-EDTA (TE) Buffer. This was used as a concentrate and aliquots were made at 100 ng/µL. A polymerase chain reaction (PCR) was performed using *P. patens* gDNA and promoter targeted primers. PCR was run multiple times with multiple annealing temperatures. PCR settings were as follows: a denaturation at 95°C for 15 s, and 35 cycles of annealing at 50°C, 53°C, 55°C, 58°C, and 61°C for 90 s, and extension at 72°C for 210s. The PCR products and reporter gene construct are separately digested with the restriction enzyme Cfr9I. Products are then run on a gel and extracted with a gel and PCR clean-up kit.

Fragments are then ligated together using a T4 ligase, and cloned into *E. coli* for plasmid extraction and long-term storage at -20°C.

### ***CRISPR/Cas9 Construct for Enhanced Transformation Efficiency***

CRISPR/Cas9 constructs were generated based on procedures performed in previous work but instead cut sites were targeted for LTR-RT/eYFP or Pp108 loci [Brooks et al. 2014]. Potential CRISPR/Cas9 gRNA were predicted by Benchling within the LTR-RT/eYFP and Pp108 loci and 2 target sequences near homologous regions of inserts but not overlapping were chosen for both. Identified gRNA was translated into gBLOCK template and ordered from Integrated DNA Technologies (IDT) [Brooks et al. 2014]. Each gBLOCK was prepared following recommended instructions provided by IDT; diluting with TE-buffer, incubating at 55°C, vortexing, centrifuging, and storing long-term at -20°C. Assembly of CRISPR/Cas9 and gBLOCKs were constructed in Golden Gate. The assembled plasmid was transformed into competent *E. coli* and confirmed with colony PCR.

### ***Homologous Recombination Transformation in P. patens (Future Directions)***

Multiple *P. patens* (Gransden 2004) lines were used for transformation depending on whether homologous recombination was targeted to the Pp108 loci (reporter gene constructs), or whether targeted to the LTR-RT/eYFP loci. For reporter genes, *P. patens* ecotype:40001 was used [Ashton & Cove 1977; Mathieu et al. 2024]. Previous CPS/KS knockout lines (pBK3), which are incapable of synthesizing *ent*-kaurenoic acid, were previously targeted for knockout at LTR-RT sites, creating a panel of LTR-RT/eYFP mutant lines [Zhan et al. 2015, King et al. 2016, Banerjee et al. 2019]. Here RT/YFP AA (high expression), RT/YFP F (high expression), RT/YFP J (moderate expression), RT/YFP V (low expression) were used as targets for secondary knockout with LTR-RT/eYFP targeted diterpene module. Transformation via homologous

recombination was performed for all cases using the PEG-mediated transformation protocol performed previously [Liu & Vidali 2011, Banerjee et al. 2019]. The respective CRISPR/Cas9 constructs were included in transformation (1 µg). These were designed for double stranded breaks at Pp108 and LTR-RT/eYFP loci for increased homologous recombination and transformation success [Collonnier et al. 2017, Zhu 2021]. Construct derived mutants and LTR-RT/eYFP derived mutants are transferred after 4 days to sulfonamide and Kanamycin selection media treatment respectively.

### ***Microscopy (Future Directions)***

Both eYFP and chlorophyll fluorescence are measured using Fluoview FV 10i (Olympus), at excitation 480 nm/emission 527 nm (eYFP) and excitation 559 nm/emissions 570/670 nm (chlorophyll), as was done previously [Banerjee et al. 2019].

### ***Validation of Successful Transformation (Future Directions)***

Initial transformation validation to confirm the presence of promoter reporter mutants and LTR-RT/YFP mutants examine the presence/absence of YFP fluorescence under blue light.

Confocal microscopy follow and serve as a more rigorous validation protocol. Depending on the transformants being investigated, eYFP expression (promoter reporter mutants) or tentative absence due to knockout (LTR-RT/eYFP mutants) are used for further confirmation and imaging.

Final confirmation validates presence with PCR and subsequent sequence-verification. For reporter genes, if eYFP expression is absent, PCR validation must also be performed as well due to the tentative and subjective expression based on promoter conditional response.

### ***Analysis with Gas-Chromatography Mass-Spectrometry (GC-MS) (Future Directions)***

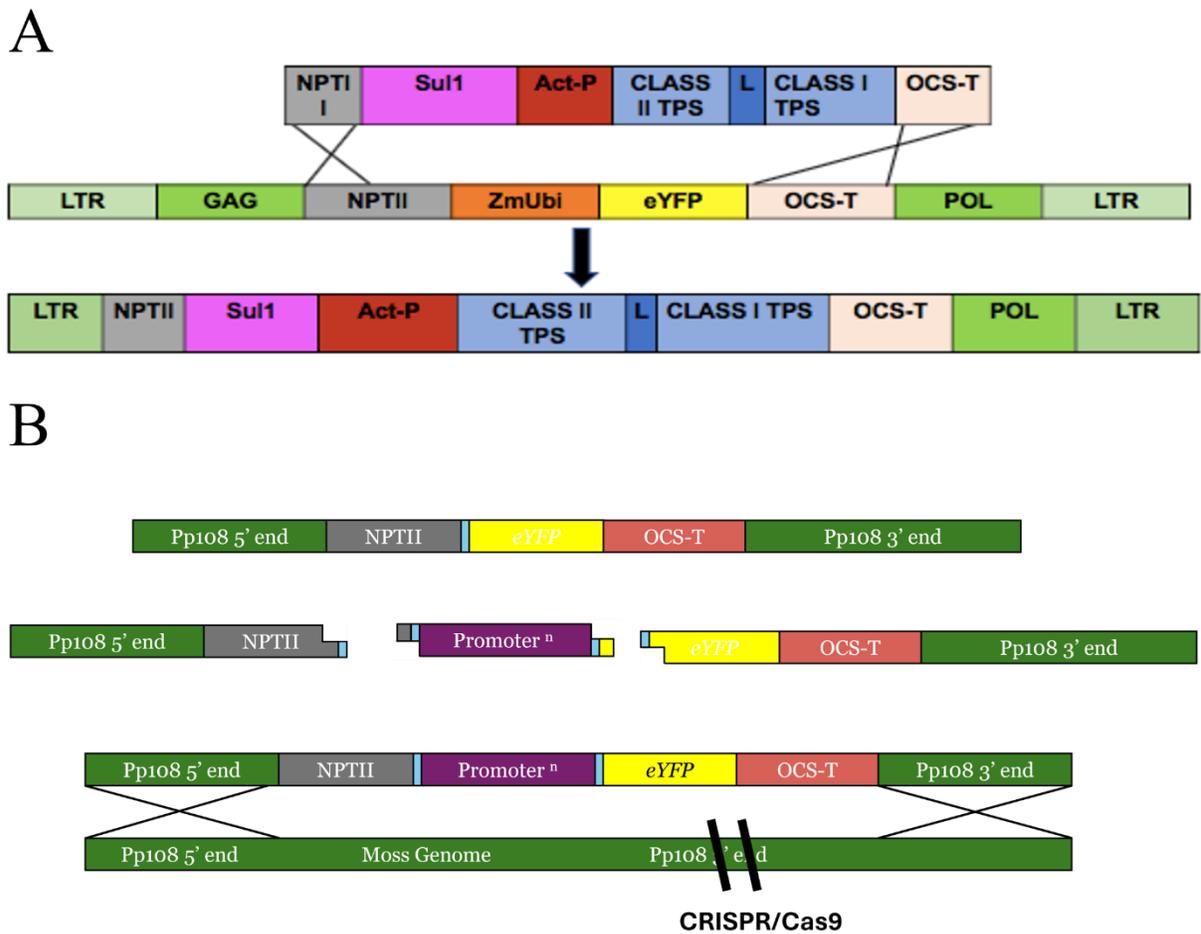
All analyses follow the performed methods and materials in previous work [Banerjee et al. 2019].

### ***Conditional Response of Specific Promoters (Future Directions)***

After sufficient recovery and validation of successful transformation, confirmed reporter/promoter lines are investigated for a response conditional to their environment. Mutants are propagated with fungal coculture (*L. elongata* or *B. erionia*) or in isolation (control) and plated on BCD agar media for two weeks, like in previous work [Mathieu et al. 2024]. *P. patens* tissue are collected and investigated with microscopy measuring eYFP and chlorophyll excitation.

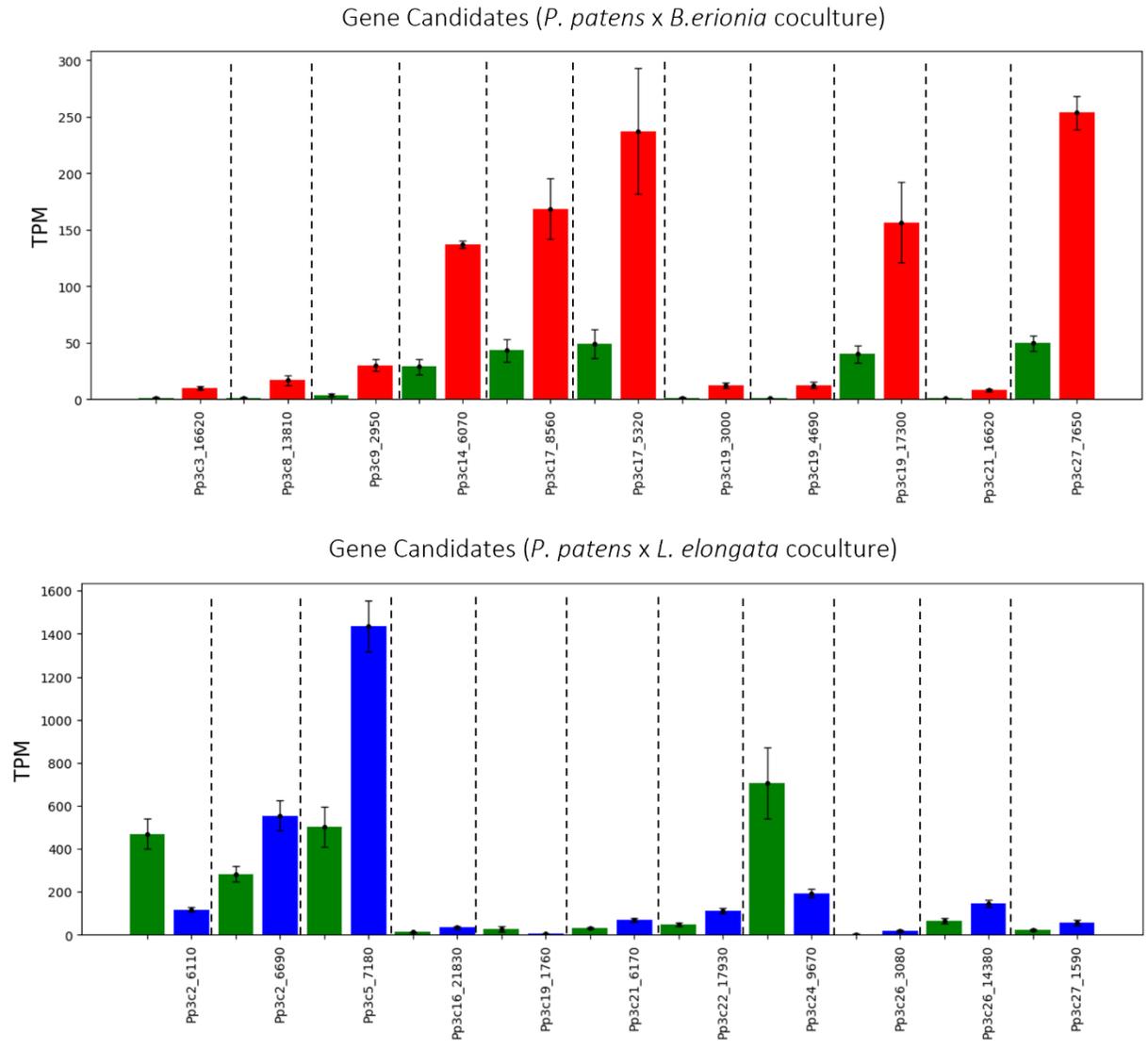
### **Results**

DNA constructs for homologous recombination were all successfully assembled and sequence verified, including the LTR-RT/eYFP diterpene module, the empty reporter gene cassette, and CRISPR/Cas9 system for targeted double stranded break at LTR-RT/eYFP and Pp108 loci (Figure 4.1). Gene candidates were selected based on gene response in the presence and absence of fungal cocultures (Figure 4.2). Promoter regions were selected based on the upstream region of candidate gene sequence which had high corresponding transcription factor binding site motifs (Figure 4.3). Many of the promoters were also successfully isolated from *P. patens* gDNA. Despite promoters being notoriously challenging for cloning, all but one *P. patens* x *B. erionia* (Pp3c21\_7650) gene could not be isolated. Necessary conditions for PCR annealing, product size, gene function, forward primer sequence, and reverse primer sequence can be found in Table 4.1.



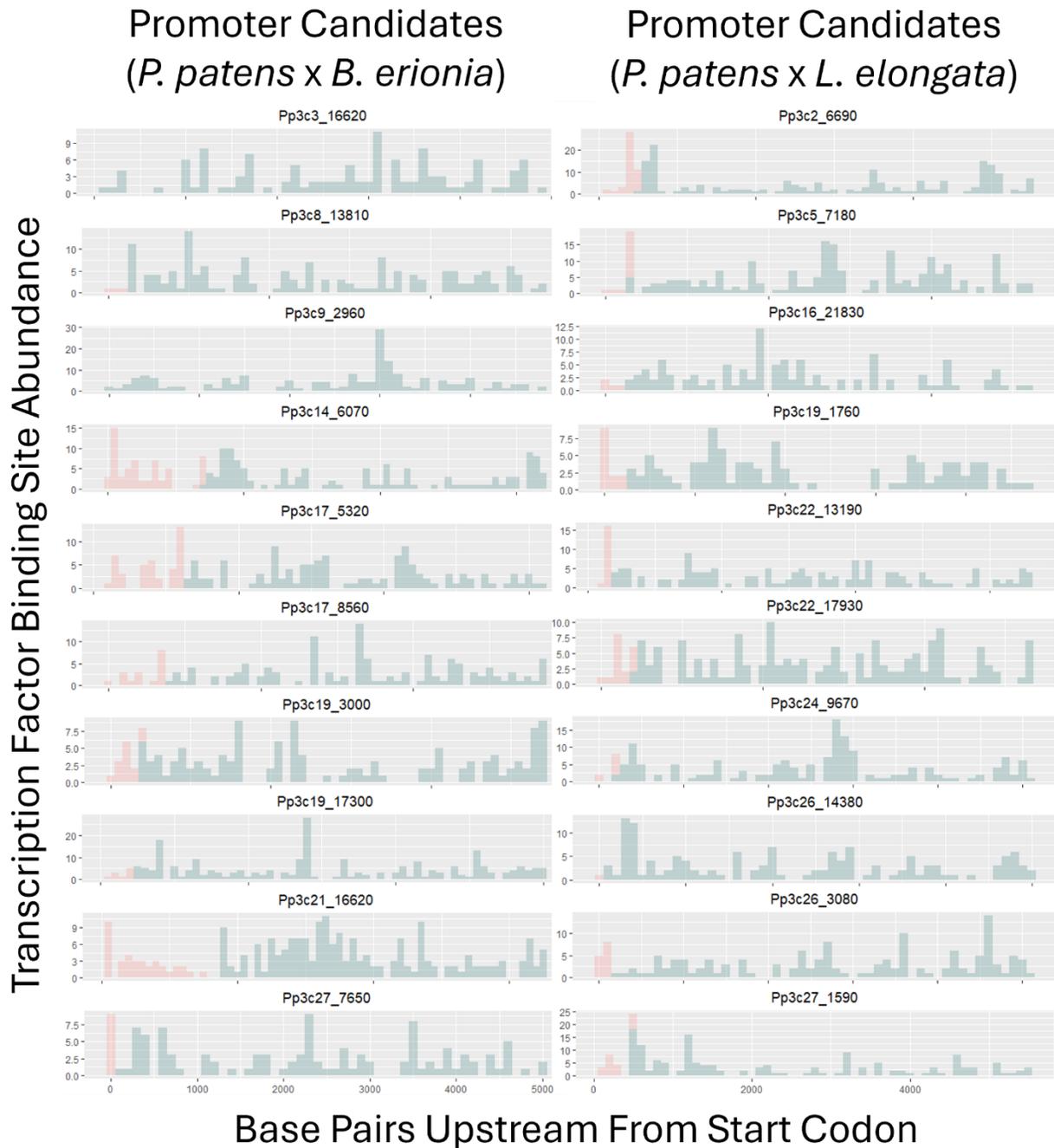
**Figure 4.1:** Homologous DNA constructs for *P. patens* transformation

DNA constructs targeted for homologous recombination with the aid of CRISPR/Cas9 constructs of the **A** LTR-RT/eYFP mutants eYFP loci with diterpene module and **B** *P. patens*, ecotype: Pp40001 strain with reporter gene construct targeted for the Pp108 loci



**Figure 4.2:** *Physcomitrium patens* transcript per million abundance of select genes grown in isolation and in coculture with either *B. erionia* or *L. elongata*

*Physcomitrium patens* expression profiles from selected gene candidates with statistical support and high responsivity to the presence of *B. erionia* coculture (Red) or *L. elongata* coculture (Blue) compared to when grown in isolation (Green).



**Figure 4.3:** Identified transcription factor binding motifs within the 5' UTR and 5kbp upstream of fungal responsive gene candidates

Predicted transcription factor binding motifs from *Physcomitrella patens* based on the results from the program PlantRegMap/PlantTFDB. Sequences investigate everything before the identified start site, where the 5'UTR is represented in pink and the 5kbp upstream of the UTR is represented as blue. Left side represents gene candidates with activity in response to *B. erionia* and the right side for genes responsive to *L. elongata*

**Table 4.1:** Candidate genes for promoter cloning with *P. patens* reporter gene construct

Gene of Interest	Coculture	Tm	bp	Gene function	Forward Primer	Reverse Primer
Pp3c3_16620	<i>B. erionia</i>	58°C	2603	early light induced protein	TACCCGGGGCCTTTTAGGCTTAGACTCTTC	TTCCCGGGTCGATGATTGAATCGAAGC
Pp3c17_13810	<i>B. erionia</i>	58°C	2033	unknown	TACCCGGGTCATTAATAGAACTCACACTG	TTCCCGGGACAACACCAAATCACTTGC
Pp3c9_2960	<i>B. erionia</i>	55°C	3072	unknown	CATGTAGCGACGACGAC	TTCCCGGGGAGCCAATATGATCGCCC
Pp3c14_6070	<i>B. erionia</i>	50°C	2211	transporter	TACCCGGGCATTATTGTCGCACCTTGTAAGG	TTCCCGGGTTCTTTCGCTTCTCTACGC
Pp3c17_5320	<i>B. erionia</i>	55°C	2004	unknown	TACCCGGGAGCAACAGAGCCATGACAACCTG	TTCCCGGGTTCTCGGTTAGTTTCGCTTCG
Pp3c17_8560	<i>B. erionia</i>	53°C	3797	abioitic stress response	CAAGTCAAGTAAAAGAAAGAACTTGAG	TTCCCGGGTTTCTGAGTATGAATGAAATGCACTC
Pp3c19_3000	<i>B. erionia</i>	55°C	2808	transcription factor	AACCCGGGCTCATGATTCTAGGAGTTC	AACCCGGGAACACCCAGAAATAAAGTGG
Pp3c19_4690	<i>B. erionia</i>	61°C	3110	syntaxin-related gene	TACCCGGGTCGGACGACTACAAGTTCAACC	TTCCCGGGGTTATCAGACTGTGAAATCGCG
Pp3c19_17300	<i>B. erionia</i>	58°C	2015	gametophore-coexpressed	TACCCGGGAGCTGTGGCAGTAATAAAACAACG	TTCCCGGGCTCTCTCACAACTCTCGTC
Pp3c21_16620	<i>B. erionia</i>		3968	carbon metabolism	CCCCCGGGTAAAGGCACAAAGGTTATCCATAC	TTCCCGGGCTTGACAACCTTCTCTGCTGCC
Pp3c27_7650	<i>B. erionia</i>	55°C	1016	membrane protein	TGTTGTTTGACTTTTCAAGGTCT	CTTCGAGACTCTCAACGTCACC
Pp3c2_6110	<i>L. elongata</i>	58°C	702	unknown	TACCCGGGCGAATTATGAACCTCACTAC	TTCCCGGGTCAATATCTATGGTTGAGTG
Pp3c2_6690	<i>L. elongata</i>	61°C	1139	transcription factor	TACCCGGGTTTATCAAAAACACATGCAATATC	TTCCCGGGCTTCGATTACCTGCAAAAAC
Pp3c16_21830	<i>L. elongata</i>	61°C	2526	unknown	TACCCGGGAAGATCCATGCTTGTGTGC	TTCCCGGGAGTTTACCAGAAGAATCGAA
Pp3c22_17930	<i>L. elongata</i>	61°C	3834	unknown	TACCCGGGCAATTAGTCTAGTCCACCGAGCC	TTCCCGGGCTACACATTGACAAAGTCTCG
Pp3c26_3080	<i>L. elongata</i>	61°C	1873	peroxidase	TACCCGGGGTCAATATTGGGGTCAATA	TTCCCGGGTGTTCCTCTCAAACAAGG
Pp3c5_7180	<i>L. elongata</i>	50°C	1564	chlorophyl A/B binding	ATCCCGGGGTACAGATCAATCCAGTTGC	TTCCCGGGGCTGAAACACACAATGCAC
Pp3c19_1760	<i>L. elongata</i>	55°C	2029	periostin	TACCCGGGTTTTATGGACATGATAGTAG	TTCCCGGGGTGAGGTGGGCAGCATAGC
Pp3c21_6170	<i>L. elongata</i>	55°C	3042	polygalacturonase; pectinase	TACCCGGGTTTGCAACAGACTTATCTGAGG	TTCCCGGGCCACACAGACAAAGCCTAC
Pp3c24_9670	<i>L. elongata</i>	55°C	2973	photosynthesis	TACCCGGGCATACGATGTAATGTCTATGG	TTCCCGGGCTTCAGAAAAACCAATATCTCTCTC
Pp3c26_14380	<i>L. elongata</i>	58°C	1526	mitosis related	TACCCGGGTCATAGCTCCACCGGTGATC	TTCCCGGGATCTATTCTCTGGCTGCATATC
Pp3c27_1590	<i>L. elongata</i>	55°C	2020	gametophore-coexpressed	TACCCGGGGTGTTCCTCCGTGCTGCTTG	GGTTCCAATAGTCGAGCT

## Conclusion & Future Directions

This project remains incomplete due to a lack of available time and challenges faced with the successful transformation of *P. patens*. If transformation had been successful, experimental procedures would have progressed for the identification of predicted diterpene production and validation of eYFP conditional expression. The design of these constructs provide a new perspective for synthetic biology in *P. patens*. By targeting unconventional loci for cloning and determining conditional and local regulatory elements allows us to improve *P. patens* as a platform for synthetic biology

## REFERENCES

- Aros, D., Gonzalez, V., Allemann, R. K., Müller, C. T., Rosati, C., & Rogers, H. J. (2012). Volatile emissions of scented *Alstroemeria* genotypes are dominated by terpenes, and a myrcene synthase gene is highly expressed in scented *Alstroemeria* flowers. *Journal of Experimental Botany*, *63*(7), 2739–2752. <https://doi.org/10.1093/jxb/err456>
- Ashton, N. W., & Cove, D. J. (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. *Molecular and General Genetics MGG*, *154*(1), 87–95. <https://doi.org/10.1007/BF00265581>
- Banerjee, A., Arnesen, J. A., Moser, D., Motsa, B. B., Johnson, S. R., & Hamberger, B. (2019). Engineering modular diterpene biosynthetic pathways in *Physcomitrella patens*. *Planta*, *249*(1), 221–233. <https://doi.org/10.1007/s00425-018-3053-0>
- Boncan, D. A. T., Tsang, S. S. K., Li, C., Lee, I. H. T., Lam, H.-M., Chan, T.-F., & Hui, J. H. L. (2020). Terpenes and Terpenoids in Plants: Interactions with Environment and Insects. *International Journal of Molecular Sciences*, *21*(19), Article 19. <https://doi.org/10.3390/ijms21197382>
- Brooks, C., Nekrasov, V., Lippman, Z. B., & Van Eck, J. (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiology*, *166*(3), 1292–1297. <https://doi.org/10.1104/pp.114.247577>
- Buckingham, J. (2015). *Natural Products Desk Reference*. CRC Press.
- Buckingham, J. (Ed.). (2023). *Dictionary of Natural Products, Supplement 2*. Routledge. <https://doi.org/10.1201/9781315141169>
- Caissard, J.-C., Meekijironroj, A., Baudino, S., & Anstett, M.-C. (2004). Localization of production and emission of pollinator attractant on whole leaves of *Chamaerops humilis* (Arecaceae). *American Journal of Botany*, *91*(8), 1190–1199. <https://doi.org/10.3732/ajb.91.8.1190>
- Celedon, J. M., & Bohlmann, J. (2016). Genomics-Based Discovery of Plant Genes for Synthetic Biology of Terpenoid Fragrances: A Case Study in Sandalwood oil Biosynthesis. *Methods in Enzymology*, *576*, 47–67. <https://doi.org/10.1016/bs.mie.2016.03.008>
- Chou, M.-Y., Andersen, T. B., Mehan Llontop, M. E., Beculheimer, N., Sow, A., Moreno, N., Shade, A., Hamberger, B., & Bonito, G. (2023). Terpenes modulate bacterial and fungal growth and sorghum rhizobiome communities. *Microbiology Spectrum*, *11*(5), e01332-23. <https://doi.org/10.1128/spectrum.01332-23>
- Chow, C.-N., Lee, T.-Y., Hung, Y.-C., Li, G.-Z., Tseng, K.-C., Liu, Y.-H., Kuo, P.-L., Zheng, H.-Q., & Chang, W.-C. (2019). PlantPAN3.0: A new and updated resource for reconstructing

- transcriptional regulatory networks from CHIP-seq experiments in plants. *Nucleic Acids Research*, 47(D1), D1155–D1163. <https://doi.org/10.1093/nar/gky1081>
- Christensen, A. H., & Quail, P. H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research*, 5(3), 213–218. <https://doi.org/10.1007/BF01969712>
- Christianson, D. W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*, 117(17), 11570–11648. <https://doi.org/10.1021/acs.chemrev.7b00287>
- Collonnier, C., Epert, A., Mara, K., Maclot, F., Guyon-Debast, A., Charlot, F., White, C., Schaefer, D. G., & Nogué, F. (2017). CRISPR-Cas9-mediated efficient directed mutagenesis and RAD51-dependent and RAD51-independent gene targeting in the moss *Physcomitrella patens*. *Plant Biotechnology Journal*, 15(1), 122–131. <https://doi.org/10.1111/pbi.12596>
- Cove, D. (2005). The Moss *Physcomitrella patens*. *Annual Review of Genetics*, 39(1), 339–358. <https://doi.org/10.1146/annurev.genet.39.073003.110214>
- Cseke, L. J., Kaufman, P. B., & Kirakosyan, A. (2007). The Biology of Essential Oils in the Pollination of Flowers. *Natural Product Communications*, 2(12), 1934578X0700201225. <https://doi.org/10.1177/1934578X0700201225>
- Degenhardt, J., Gershenzon, J., Baldwin, I. T., & Kessler, A. (2003). Attracting friends to feast on foes: Engineering terpene emission to make crop plants more attractive to herbivore enemies. *Current Opinion in Biotechnology*, 14(2), 169–176. [https://doi.org/10.1016/S0958-1669\(03\)00025-9](https://doi.org/10.1016/S0958-1669(03)00025-9)
- Dötterl, S., & Gershenzon, J. (2023). Chemistry, biosynthesis and biology of floral volatiles: Roles in pollination and other functions. *Natural Product Reports*, 40(12), 1901–1937. <https://doi.org/10.1039/D3NP00024A>
- Erbilgin, N., Krokene, P., Christiansen, E., Zeneli, G., & Gershenzon, J. (2006). Exogenous application of methyl jasmonate elicits defenses in Norway spruce (*Picea abies*) and reduces host colonization by the bark beetle *Ips typographus*. *Oecologia*, 148(3), 426–436. <https://doi.org/10.1007/s00442-006-0394-3>
- Gershenzon, J., & Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nature Chemical Biology*, 3(7), Article 7. <https://doi.org/10.1038/nchembio.2007.5>
- González-Coloma, A., Guadaño, A., Tonn, C. E., & Sosa, M. E. (2005). Antifeedant/Insecticidal Terpenes from Asteraceae and Labiatae Species Native to Argentinean Semi-arid Lands. *Zeitschrift Für Naturforschung C*, 60(11–12), 855–861. <https://doi.org/10.1515/znc-2005-11-1207>

- Hausch, B. J., Lorjaroenphon, Y., & Cadwallader, K. R. (2015). Flavor chemistry of lemon-lime carbonated beverages. *Journal of Agricultural and Food Chemistry*, 63(1), 112–119. <https://doi.org/10.1021/jf504852z>
- Hayashi, K., Horie, K., Hiwatashi, Y., Kawaide, H., Yamaguchi, S., Hanada, A., Nakashima, T., Nakajima, M., Mander, L. N., Yamane, H., Hasebe, M., & Nozaki, H. (2010). Endogenous Diterpenes Derived from *ent*-Kaurene, a Common Gibberellin Precursor, Regulate Protonema Differentiation of the Moss *Physcomitrella patens*. *Plant Physiology*, 153(3), 1085–1097. <https://doi.org/10.1104/pp.110.157909>
- Heiling, S., Schuman, M. C., Schoettner, M., Mukerjee, P., Berger, B., Schneider, B., Jassbi, A. R., & Baldwin, I. T. (2010). Jasmonate and ppHsystemin Regulate Key Malonylation Steps in the Biosynthesis of 17-Hydroxygeranylinalool Diterpene Glycosides, an Abundant and Effective Direct Defense against Herbivores in *Nicotiana attenuata*. *The Plant Cell*, 22(1), 273–292. <https://doi.org/10.1105/tpc.109.071449>
- Heusler, F. (1902). *The Chemistry of the Terpenes*. P. Blakiston's son & Company.
- Hoffmann, B., Proust, H., Belcram, K., Labrune, C., Boyer, F.-D., Rameau, C., & Bonhomme, S. (2014). Strigolactones Inhibit Caulonema Elongation and Cell Division in the Moss *Physcomitrella patens*. *PLOS ONE*, 9(6), e99206. <https://doi.org/10.1371/journal.pone.0099206>
- Huang, A. C., & Osbourn, A. (2019). Plant terpenes that mediate below-ground interactions: Prospects for bioengineering terpenoids for plant protection. *Pest Management Science*, 75(9), 2368–2377. <https://doi.org/10.1002/ps.5410>
- Ikram, N. K. K., Zakariya, A. M., Saiman, M. Z., Kashkooli, A. B., & Simonsen, H. T. (2023). Heterologous Production of Artemisinin in *Physcomitrium patens* by Direct in vivo Assembly of Multiple DNA Fragments. *Bio-Protocol*, 13(14), e4719. <https://doi.org/10.21769/BioProtoc.4719>
- Jassbi, A. R., Gase, K., Hettenhausen, C., Schmidt, A., & Baldwin, I. T. (2008). Silencing Geranylgeranyl Diphosphate Synthase in *Nicotiana attenuata* Dramatically Impairs Resistance to Tobacco Hornworm. *Plant Physiology*, 146(3), 974–986. <https://doi.org/10.1104/pp.107.108811>
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 45(D1), D1040–D1045. <https://doi.org/10.1093/nar/gkw982>
- Johnson, S. R., Bhat, W. W., Bibik, J., Turmo, A., Hamberger, B., Evolutionary Mint Genomics Consortium, null, & Hamberger, B. (2019). A database-driven approach identifies additional diterpene synthase activities in the mint family (Lamiaceae). *The Journal of Biological Chemistry*, 294(4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>

- Kawai, Y., Ono, E., & Mizutani, M. (2014). Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants. *The Plant Journal*, 78(2), 328–343. <https://doi.org/10.1111/tpj.12479>
- Keeling, C. I., & Bohlmann, J. (2006). Diterpene resin acids in conifers. *Phytochemistry*, 67(22), 2415–2423. <https://doi.org/10.1016/j.phytochem.2006.08.019>
- King, B. C., Vavitsas, K., Ikram, N. K. B. K., Schröder, J., Scharff, L. B., Bassard, J.-É., Hamberger, B., Jensen, P. E., & Simonsen, H. T. (2016). In vivo assembly of DNA-fragments in the moss, *Physcomitrella patens*. *Scientific Reports*, 6(1), 25030. <https://doi.org/10.1038/srep25030>
- Koul, O. (2008). Phytochemicals and Insect Control: An Antifeedant Approach. *Critical Reviews in Plant Sciences*, 27(1), 1–24. <https://doi.org/10.1080/07352680802053908>
- Kutyna, D. R., & Borneman, A. R. (2018). Heterologous Production of Flavour and Aroma Compounds in *Saccharomyces cerevisiae*. *Genes*, 9(7), Article 7. <https://doi.org/10.3390/genes9070326>
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R., Vives, C., Morata, J., Symeonidi, A., Hiss, M., Muchero, W., Kamisugi, Y., Saleh, O., Blanc, G., Decker, E. L., ... Rensing, S. A. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal*, 93(3), 515–533. <https://doi.org/10.1111/tpj.13801>
- Lange, B. M., Mahmoud, S. S., Wildung, M. R., Turner, G. W., Davis, E. M., Lange, I., Baker, R. C., Boydston, R. A., & Croteau, R. B. (2011). Improving peppermint essential oil yield and composition by metabolic engineering. *Proceedings of the National Academy of Sciences*, 108(41), 16944–16949. <https://doi.org/10.1073/pnas.1111558108>
- Laurent, P., Braekman, J.-C., Daloz, D., & Pasteels, J. (2003). Biosynthesis of Defensive Compounds from Beetles and Ants. *European Journal of Organic Chemistry*, 2003(15), 2733–2743. <https://doi.org/10.1002/ejoc.200300008>
- Li, J., Halitschke, R., Li, D., Paetz, C., Su, H., Heiling, S., Xu, S., & Baldwin, I. T. (2021). Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science*, 371(6526), 255–260. <https://doi.org/10.1126/science.abe4713>
- Lipińska, M. M., Gołębiowski, M., Szlachetko, D. L., & Kowalkowska, A. K. (2022). Floral attractants in the black orchid *Brasiliorchis schunkeana* (Orchidaceae, Maxillariinae): Clues for presumed sapromyophily and potential antimicrobial activity. *BMC Plant Biology*, 22(1), 575. <https://doi.org/10.1186/s12870-022-03944-8>
- Liu, Y.-C., & Vidali, L. (2011). Efficient Polyethylene Glycol (PEG) Mediated Transformation of the Moss *Physcomitrella patens*. *JoVE (Journal of Visualized Experiments)*, 50, e2560. <https://doi.org/10.3791/2560>

- Lu, X., Zhang, J., Brown, B., Li, R., Rodríguez-Romero, J., Berasategui, A., Liu, B., Xu, M., Luo, D., Pan, Z., Baerson, S. R., Gershenzon, J., Li, Z., Sesma, A., Yang, B., & Peters, R. J. (2018). Inferring Roles in Defense from Metabolic Allocation of Rice Diterpenoids. *The Plant Cell*, *30*(5), 1119–1131. <https://doi.org/10.1105/tpc.18.00205>
- Miller, G. P., Bhat, W. W., Lanier, E. R., Johnson, S. R., Mathieu, D. T., & Hamberger, B. (2020). The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all-cis prenyl intermediate. *The Plant Journal*, *104*(3), 693–705. <https://doi.org/10.1111/tpj.14957>
- Miyazaki, S., Nakajima, M., & Kawaide, H. (2015). Hormonal diterpenoids derived from *ent*-kaurenoic acid are involved in the blue-light avoidance response of *Physcomitrella patens*. *Plant Signaling & Behavior*, *10*(2), e989046. <https://doi.org/10.4161/15592324.2014.989046>
- Nagel, R., Berasategui, A., Paetz, C., Gershenzon, J., & Schmidt, A. (2014). Overexpression of an Isoprenyl Diphosphate Synthase in Spruce Leads to Unexpected Terpene Diversion Products That Function in Plant Defense. *Plant Physiology*, *164*(2), 555–569. <https://doi.org/10.1104/pp.113.228940>
- Ndi, C. P., Semple, S. J., Griesser, H. J., Pyke, S. M., & Barton, M. D. (2007). Antimicrobial compounds from the Australian desert plant *Eremophila neglecta*. *Journal of Natural Products*, *70*(9), 1439–1443. <https://doi.org/10.1021/np070180r>
- Novikova, O., Mayorov, V., Smyshlyaev, G., Fursov, M., Adkison, L., Pisarenko, O., & Blinov, A. (2008). Novel clades of chromodomain-containing Gypsy LTR retrotransposons from mosses (Bryophyta). *The Plant Journal*, *56*(4), 562–574. <https://doi.org/10.1111/j.1365-313X.2008.03621.x>
- Nuutinen, T. (2018). Medicinal properties of terpenes found in *Cannabis sativa* and *Humulus lupulus*. *European Journal of Medicinal Chemistry*, *157*, 198–228. <https://doi.org/10.1016/j.ejmech.2018.07.076>
- Pateraki, I., Andersen-Ranberg, J., Hamberger, B., Heskes, A. M., Martens, H. J., Zerbe, P., Bach, S. S., Møller, B. L., Bohlmann, J., & Hamberger, B. (2014). Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolol, Is Synthesized in Specialized Root Cork Cells in *Coleus forskohlii*. *Plant Physiology*, *164*(3), 1222–1236. <https://doi.org/10.1104/pp.113.228429>
- Philippe, R. N., De Mey, M., Anderson, J., & Ajikumar, P. K. (2014). Biotechnological production of natural zero-calorie sweeteners. *Current Opinion in Biotechnology*, *26*, 155–161. <https://doi.org/10.1016/j.copbio.2014.01.004>
- Piccoli, P., & Bottini, R. (2013). Terpene Production by Bacteria and its Involvement in Plant Growth Promotion, Stress Alleviation, and Yield Increase. In *Molecular Microbial Ecology*

*of the Rhizosphere* (pp. 335–343). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781118297674.ch31>

- Proffit, M., Lapeyre, B., Buatois, B., Deng, X., Arnal, P., Gouzerh, F., Carrasco, D., & Hossaert-McKey, M. (2020). Chemical signal is in the blend: Bases of plant-pollinator encounter in a highly specialized interaction. *Scientific Reports*, *10*(1), Article 1.  
<https://doi.org/10.1038/s41598-020-66655-w>
- Sainsbury, F., Thuenemann, E. C., & Lomonossoff, G. P. (2009). pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal*, *7*(7), 682–693. <https://doi.org/10.1111/j.1467-7652.2009.00434.x>
- Schaefer, D. G., & Zrýd, J. P. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *The Plant Journal: For Cell and Molecular Biology*, *11*(6), 1195–1206.  
<https://doi.org/10.1046/j.1365-313x.1997.11061195.x>
- Schalk, M., Pastore, L., Mirata, M. A., Khim, S., Schouwey, M., Deguerry, F., Pineda, V., Rocci, L., & Daviet, L. (2012). Toward a Biosynthetic Route to Sclareol and Amber Odorants. *Journal of the American Chemical Society*, *134*(46), 18900–18903.  
<https://doi.org/10.1021/ja307404u>
- Schiebe, C., Hammerbacher, A., Birgersson, G., Witzell, J., Brodelius, P. E., Gershenzon, J., Hansson, B. S., Krokene, P., & Schlyter, F. (2012). Inducibility of chemical defenses in Norway spruce bark is correlated with unsuccessful mass attacks by the spruce bark beetle. *Oecologia*, *170*(1), 183–198. <https://doi.org/10.1007/s00442-012-2298-8>
- Schulte, J., & Reski, R. (2004). High Throughput Cryopreservation of 140 000 *Physcomitrella patens* Mutants. *Plant Biology*, *6*(2), 119–127. <https://doi.org/10.1055/s-2004-817796>
- Tetali, S. D. (2019). Terpenes and isoprenoids: A wealth of compounds for global use. *Planta*, *249*(1), 1–8. <https://doi.org/10.1007/s00425-018-3056-x>
- Theis, N., & Lerchau, M. (2003). The Evolution of Function in Plant Secondary Metabolites. *International Journal of Plant Sciences*, *164*(S3), S93–S102.  
<https://doi.org/10.1086/374190>
- Toyomasu, T., Usui, M., Sugawara, C., Otomo, K., Hirose, Y., Miyao, A., Hirochika, H., Okada, K., Shimizu, T., Koga, J., Hasegawa, M., Chuba, M., Kawana, Y., Kuroda, M., Minami, E., Mitsuhashi, W., & Yamane, H. (2014). Reverse-genetic approach to verify physiological roles of rice phytoalexins: Characterization of a knockdown mutant of OsCPS4 phytoalexin biosynthetic gene in rice. *Physiologia Plantarum*, *150*(1), 55–62.  
<https://doi.org/10.1111/ppl.12066>
- Vendrell-Mir, P., López-Obando, M., Nogué, F., & Casacuberta, J. M. (2020). Different Families of Retrotransposons and DNA Transposons Are Actively Transcribed and May Have

Transposed Recently in *Physcomitrium (Physcomitrella) patens*. *Frontiers in Plant Science*, 11. <https://doi.org/10.3389/fpls.2020.01274>

- Wang, G., Tang, W., & Bidigare, R. R. (2005). Terpenoids As Therapeutic Drugs and Pharmaceutical Agents. In L. Zhang & A. L. Demain (Eds.), *Natural Products: Drug Discovery and Therapeutic Medicine* (pp. 197–227). Humana Press. [https://doi.org/10.1007/978-1-59259-976-9\\_9](https://doi.org/10.1007/978-1-59259-976-9_9)
- Wang, Z., R. Nelson, D., Zhang, J., Wan, X., & J. Peters, R. (2023). Plant (di)terpenoid evolution: From pigments to hormones and beyond. *Natural Product Reports*, 40(2), 452–469. <https://doi.org/10.1039/D2NP00054G>
- Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, 93(9), 2325–2327. <https://doi.org/10.1021/ja00738a045>
- Wilson, S. A., & Roberts, S. C. (2012). Recent advances towards development and commercialization of plant cell culture processes for the synthesis of biomolecules. *Plant Biotechnology Journal*, 10(3), 249–268. <https://doi.org/10.1111/j.1467-7652.2011.00664.x>
- Zeng, T., Liu, Z., Zhuang, J., Jiang, Y., He, W., Diao, H., Lv, N., Jian, Y., Liang, D., Qiu, Y., Zhang, R., Zhang, F., Tang, X., & Wu, R. (2020). TeroKit: A Database-Driven Web Server for Terpenome Research. *Journal of Chemical Information and Modeling*, 60(4), 2082–2090. <https://doi.org/10.1021/acs.jcim.0c00141>
- Zerbe, P., & Bohlmann, J. (2014). Bioproducts, Biofuels, and Perfumes: Conifer Terpene Synthases and their Potential for Metabolic Engineering. In R. Jetter (Ed.), *Phytochemicals – Biosynthesis, Function and Application: Volume 44* (pp. 85–107). Springer International Publishing. [https://doi.org/10.1007/978-3-319-04045-5\\_5](https://doi.org/10.1007/978-3-319-04045-5_5)
- Zerbe, P., Chiang, A., Yuen, M., Hamberger, B., Hamberger, B., Draper, J. A., Britton, R., & Bohlmann, J. (2012). Bifunctional cis-Abienol Synthase from *Abies balsamea* Discovered by Transcriptome Sequencing and Its Implications for Diterpenoid Fragrance Production. *The Journal of Biological Chemistry*, 287(15), 12121–12131. <https://doi.org/10.1074/jbc.M111.317669>
- Zerbe, P., Hamberger, B., Yuen, M. M. S., Chiang, A., Sandhu, H. K., Madilao, L. L., Nguyen, A., Hamberger, B., Bach, S. S., & Bohlmann, J. (2013). Gene Discovery of Modular Diterpene Metabolism in Nonmodel Systems. *Plant Physiology*, 162(2), 1073–1091. <https://doi.org/10.1104/pp.113.218347>
- Zhan, X., Bach, S. S., Hansen, N. L., Lunde, C., & Simonsen, H. T. (2015). Additional diterpenes from *Physcomitrella patens* synthesized by copalyl diphosphate/kaurene synthase (PpCPS/KS). *Plant Physiology and Biochemistry*, 96, 110–114. <https://doi.org/10.1016/j.plaphy.2015.07.011>

- Zhan, X., Zhang, Y.-H., Chen, D.-F., & Simonsen, H. T. (2014). Metabolic engineering of the moss *Physcomitrella patens* to produce the sesquiterpenoids patchoulol and  $\alpha/\beta$ -santalene. *Frontiers in Plant Science*, 5. <https://www.frontiersin.org/articles/10.3389/fpls.2014.00636>
- Zhao, D.-D., Jiang, L.-L., Li, H.-Y., Yan, P.-F., & Zhang, Y.-L. (2016). Chemical Components and Pharmacological Activities of Terpene Natural Products from the Genus *Paeonia*. *Molecules (Basel, Switzerland)*, 21(10), 1362. <https://doi.org/10.3390/molecules21101362>
- Zhao, T., Krokene, P., Hu, J., Christiansen, E., Björklund, N., Långström, B., Solheim, H., & Borg-Karlson, A.-K. (2011). Induced Terpene Accumulation in Norway Spruce Inhibits Bark Beetle Colonization in a Dose-Dependent Manner. *PLOS ONE*, 6(10), e26649. <https://doi.org/10.1371/journal.pone.0026649>
- Zhou, F., & Pichersky, E. (2020). More is better: The diversity of terpene metabolism in plants. *Current Opinion in Plant Biology*, 55, 1–10. <https://doi.org/10.1016/j.pbi.2020.01.005>
- Zhu, L. (2021). Targeted Gene Knockouts by Protoplast Transformation in the Moss *Physcomitrella patens*. *Frontiers in Genome Editing*, 3. <https://doi.org/10.3389/fgeed.2021.719087>

## APPENDIX: *CURRICULUM VITAE*

### Contact Information

Name: Davis Mathieu

Email: mathieud@msu.edu

LinkedIn: davis-mathieu

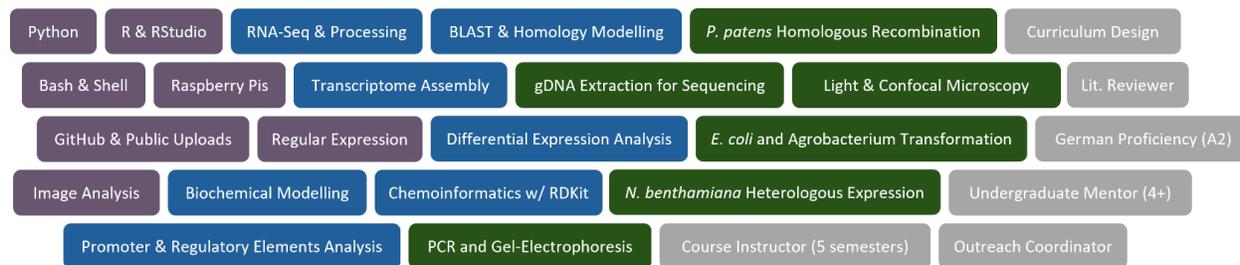
### Personal Statement: *PhD in Genetics and Genome Sciences*

---

Collaborative bioinformatician, plant geneticist, and science communicator. Leading projects with extensive breadth coordinating experts in mycology, plant biology, evolution, genomics, specialized metabolism, bioinformatics, and computational modelling. Major accomplishments include 1.) Analyzing multifaceted interaction of moss and fungi with implications into long-standing symbiosis 2.) Uncovering of complex patterns through deconstruction and reconstruction of the TeroKit database, which has over 160K unique terpene entries and 3.) The exhibition of “Fog of Dawn” presented at Science Gallery Detroit, reaching an audience of 400K people.

### Areas of Expertise

---



### Education

---

(Sept 2017 – Apr 2024, Michigan State University)

- **Doctor of Philosophy: Genetics & Genome Sciences**

(Aug 2013 – May 2017 South Dakota School of Mines & Technology)

- **Bachelor of Science: Applied Biological Sciences; Minor: Chemistry**

### Research Experience

---

**Dissertation Research (2017-2024):**

Title: ***PHYSCOMITRIUM PATENS: APPLICATIONS IN SYNTHETIC BIOLOGY AND THE CURATION OF THE DITERPENOID LIBRARIES***

1. ***Physcomitrium patens: A Chassis for Diterpene Synthesis and the Exploration of Fungal Symbiosis for Improved Growth and Extraction***
2. ***Multilevel Analysis between Physcomitrium patens and Mortierellaceae Endophytes Explores Potential Long-Standing Interaction among Land Plants and Fungi***
3. ***Rule-Based Deconstruction and Reconstruction of the Diterpene Library: A Simulation of Synthesis and Unravelling of Compound Structural Diversity***
4. ***Long Terminal Repeat Retrotransposon Targeted Transformation and Development of Promoter Reporter System in Physcomitrium patens for Sequential Targeting of Diterpene Module***

Advisor: **Dr. Björn Hamberger**

Committee: Dr. Gregory Bonito, Dr. Frances Trails, Dr. Ning Jiang, Dr. Robert Van Buren

**NEXTplant/iGRAD Exchange at Heinrich Heine Universität, Düsseldorf DE (Fall 2022):**

Project Title: ***Predicting Promiscuity & Modeling the Diterpenoid Synthesis Landscape***

Advisor: **Dr. Oliver Ebenhöh & Dr. Björn Hamberger**

#### **MSU Lab Rotations:**

Project Title: ***Preparation & Genome Assembly*** (Winter 2017)

Advisor: **Dr. C. Robin Buell**

Project Title: ***Analyzing & Identifying Rice Genome Architectural Anomalies*** (Fall 2017)

Advisor: **Dr. Ning Jiang**

#### **Undergraduate Lab Experience:**

Project Title: ***Utilization of CRISPR/Cas9 to Induce Gene knockout in the Noxious Weed species Euphorbia lathyris to Develop as a Novel Arid Biofuel Crop*** (Summer 2016)

Advisor: **Dr. Björn Hamberger**; Great Lakes Bioenergy Research Center Research Experience for

Undergraduates (GLBRC REU)

Project Title: ***Sampling and Analyzing virulence in South Dakota Watershed*** (2014-2016)

Advisor: **Kelsey Murray & Dr. Linda DeVeaux**

#### **Publications**

---

**Davis Mathieu**, Nicholas Schlecht, Marvin Van Aalst, Kevin M. Shebek, Luke Busta, Nicole Babineau, Oliver Ebenhöh, Björn Hamberger. ‘*Rule-Based Deconstruction and Reconstruction of the Diterpene Library: A Simulation of Synthesis and Unravelling of Compound Structural Diversity.*’ (IN PROGRESS).

**Davis Mathieu**, Abigail E. Bryson, Britta Hamberger, Vasanth Singan, Keykhosrow Keymanesh, Mei Wang, Kerrie Barry, Stephen Mondo, Jasmyn Pangilinan, Maxim Koriabine, Igor V. Grigoriev, Gregory Bonito, Björn Hamberger. ‘*Multilevel analysis between Physcomitrium*

*patens* and Mortierellaceae endophytes explores potential long-standing interaction among land plants and fungi.’ *The Plant Journal* **118**, 304-323 (2024).

Connor Yeck (Interviewed Björn Hamberger & **Davis Mathieu**). ‘Friend or foe? MSU researchers explore ancient partnership between moss and fungi.’ MSU NatSci & EurekaAlert. <https://natsci.msu.edu/news/2024-02-msu-researchers-explore-ancient-partnership-between-moss-and-fungi%20%20%20.aspx>; <https://www.eurekaalert.org/news-releases/1033809> (2024).

Jyothi Kumar\*, Fabio Gomez-Cano†, Seth W. Hunt†, Serena G. Lotreck†, **Davis T. Mathieu**†, McKena L. Wilson†, Tammy M. Long\*, ‘Central Dogma, Dictionaries, and Functions: Using Programming Concepts to Simulate Biological Processes.’ *CourseSource* **10** <https://doi.org/10.24918/cs.2023.24> (2023).

Abigail E. Bryson, Emily R. Lanier, Kin H. Lau, John P. Hamilton, Brienne Vaillancourt, **Davis Mathieu**, Alan E. Yoca, Garret P. Miller, Patrick P. Edger, C. Robin Buell & Björn Hamberger. ‘Uncovering a miltiradiene biosynthetic gene cluster in the Lamiaceae reaveals a dynamic evolutionary trajectory.’ *Nature Commun* **14**. <https://doi.org/10.1038/s41467-023-35845-1> (2023).

Garret P Miller, Wajid Waheed Bhat, Emily R Lanier, Sean R Johnson, **Davis T. Mathieu**, Björn Hamberger. ‘The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all-cis prenyl intermediate.’ *The Plant Journal*. <https://doi.org/10.1111/tpj.14957> (2020).\

Abigail E. Bryson, Maya Wilson Brown, Joey Mullins, Wei Dong, Keivan Bahmani, Nolan Bornowski, Christina Chiu, Philip Engelgau, Bethany Gettings, Fabio Gomezcano, Luke M. Gregory, Anna C. Haber, Donghee Hoh, Emily E. Jennings, Zhongjie Ji, Prabhjot Kaur, Sunil K. Kenchanmane Raju, Yunfei Long, Serena G. Lotreck, **Davis T. Mathieu**, Thilanka Ranaweera, Eleanore J. Ritter, Rie Sadohara, Robert Z. Shrote, Kaila E. Smith, Scott J. Teresi, Julian Venegas, Hao Wang, McKena L. Wilson, Alyssa R. Tarrant, Margaret H. Frank, Zoë Migicovsky, Jyothi Kumar, Robert VanBuren, Jason P. Londo, Daniel H. Chitwood. ‘Composite modeling of leaf shape along shoots discriminates *Vitis* species better than individual leaves.’ *Applications in Plant Sciences*. <https://doi.org/10.1002/aps3.11404> (2020).

**Davis Mathieu**. ‘*Spirit Molecules*.’ MSU Today. <https://msutoday.msu.edu/news/2020/davis-mathieu-spirit-molecules> (2020).

## Presentations

---

**Davis Mathieu**. ‘Physcomitrium patens: Applications in Synthetic Biology and the Curation of the Diterpenoid Libraries.’ Dissertation Defense (March 29<sup>th</sup>, 2024)

**Davis Mathieu**. ‘*A reconstructive and deconstructive approach for unravelling the complete diterpene library*’. Phytochemical Society of North America Annual Conference. (July 2023)

**Davis Mathieu**. ‘*Predicting Promiscuity & Modeling the Diterpenoid Synthesis Landscape*.’ Next Plant Fellows Seminar. (February 2023)

**Davis Mathieu**, Abigail Bryson, Björn Hamberger. ‘*High Throughput Phenomic Analysis of Physcomitrella patens*.’ ASBMB symposium: Evolution and Core Processes in Gene Expression. (May 2019)

**Davis Mathieu**, Abigail Bryson, Björn Hamberger. ‘*High Throughput Phenomic Analysis of Physcomitrella patens*.’ ASPB Annual Meeting (August 2019)

**Davis Mathieu**, Abigail Bryson, Björn Hamberger. ‘*High Throughput Phenomic Analysis of Physcomitrella patens*.’ MSUs Genetics Mini symposium: Epigenetic Stress Memory. (May 2019)

**Davis Mathieu**, Björn Hamberger. ‘*Science Gallery Detroit: Collisions Between Science and Art – “Fog of Dawn”*.’ Science Cafes and Pubs. (April 7<sup>th</sup>, 2019)

**Davis Mathieu**, Mitch Roth, Levi Bauer, Taylor Murphey, Christine Ponnampalam, Abby Bryson. ‘How to Catch a Criminal with DNA’. MSU Science Festival: East Lansing, MI, (April 6<sup>th</sup>, 2019)

**Davis Mathieu**, Levi Bauer, Anne-Sophie Bohrer, Laura Harding. ‘DNA has Never Been Sweeter’. Girls Math & Science Day: East Lansing, MI, (March 2<sup>nd</sup>, 2019)

**Davis Mathieu**, Hamberger, Britta; Hamberger, Björn. ‘*Transformation of Euphorbia lathyris from a Nuisance to a Biofuel*.’ MSU Mid-SURE Poster Session; East Lansing, MI, (July 2016)

Liber, Julian; Caldewell, Sarah; Lee, Jordan; Schultz, Jessica; Viola, Sophia; Uhelski, Erin; Gate, Casper; Rose, Ashley; Ducat, Danny; Hamberger, Björn; Whitehead, Tim; TerAvest, Machaela; **Davis Mathieu**; Pedro Beschoren de Costa. ‘*Climate change conjures a host of problems. Our Solution? ENDOPHYTE CLUB*.’ iGEM 2018 Jamboree; Boston, MA, October 2018

## Awards & Grants

---

- **Jeff Schell Fellowship for Agricultural Science**. The Bayer Foundation (2023).
- **NSF Research Trainee Integrated Training Model in Plant And Compu-Tational Sciences (NRT-IMPACTS) Fellow**. National Science Foundation & Michigan State University (2019-2024).

- **Plant Biotechnology for Health and Sustainability Graduate Training Program Fellow.** Michigan State University (2018).
- **RMAC Academic All American.** (Sep 2014 – May 2017)
- **Dean’s List.** South Dakota School of Mine & Technology (*Cum laude*)

### Coordinated Events

---

- **Science Gallery Detroit ‘DEPTH’ Exhibitionist: “Fog of Dawn”** (Summer 2019 – 400K audience)
- **Eli and Edythe Broad Art Museum Exhibitionist: “Spirit Molecule”** (June 2019 – September 2019)
- **Genetics and Genome Science Minisymposium Series** (Sept 17<sup>th</sup>, Sept 24<sup>th</sup>, Oct 1<sup>st</sup>, Oct 8<sup>th</sup> 2020)
- **Lansing Elementary Science STEAM Nights** (8 total events 2018-2020)
- **MSU SciComm Live Science-Art Show** (Oct 19<sup>th</sup>, 2019)
- **MSU Fascination of Plants Day** (July 2019)
- **MSU Science Festival** (March 2019)
- **Girls Math & Science Day** (March 2019)
- **MSU Fascination of Plants Day** (July 2018)
- **MSU Grandparents University** (June 2018)
- **MSU Science Festival** (March 2018)

### Select Memberships

---

- **Genetics and Genome Science Graduate Student Organization (GGS GSO)**
  - **GGS President** (May 2021 – May 2022)
  - **GGS Outreach Coordinator** (May 2018 – May 2022)
- **NRT IMPACTS Fellow** (2019 – 2024)
- **Scientific Literature Reviewer**
  - Fungal Biology* (2024 – present)
  - The Plant Cell* (2023 – present)
  - PLOS ONE* (2020 – present)
- **Team Coordinator (Great Lakes Relay/Michigan Outback Relay)** (2018 – present)

### Interests

---

Running • Kayaking • Climbing • Chess • Cooking • Geocaching • Ceramics

