"Exploring cell-type regulatory dynamics of CAM and C_4 photosynthesis in *Portulaca*"

Overview

C_4 and CAM photosynthesis are carbon concentrating mechanisms (CCMs) that have evolved as plant responses to the low CO_2 world of the past 30 million years. Both CCMs have co-opted the same set of ancient metabolic modules to boost the concentration of CO_2 needed for photosynthesis, but have deployed these modules in contrasting ways. C_4 concentrates CO_2 spatially through a two-cell CO_2 pump, while CAM accomplishes CO_2 concentration with temporally coordinated carbon storage and re-release. These adaptations confer C_4 species with the highest rates of plant photosynthesis, characterized by maize and sugarcane, and CAM plants with the highest water use efficiencies, emblematic of cacti, aloes, and agaves. Despite hundreds of independent CCM evolutions, and their shared biochemical components, only two land plant lineages are known to use both C_4 and CAM (C_4+CAM): *Portulaca* and *Trianthema*, C_4 plants that facultatively exhibit CAM in response to abiotic stress. *Portulaca*, with multiple independent origins of C_4+CAM, offers unique insights into how multiple CCMs can be integrated to increase the drought tolerance of highly productive C_4 crops. ***This research project will leverage systems and computational biology to identify the genetic elements controlling the temporal and spatial coordination of CAM and C_4 photosynthesis in Portulaca at the cell-type level***. First, I will capture expression dynamics of individual cells using single cell RNAseq and identify CCM-related *cis*-regulatory elements using assay for transposase-accessible chromatin using sequencing (ATACseq). I will construct gene regulatory networks using single cell RNAseq data and use machine learning approaches to distinguish *cis*-elements and regulatory dymanics governing C_4 and CAM. Finally, I will compare regulatory networks to identify shared and unique elements underlying the evolution of CCMs in *Portulaca*. Through the PRFB, I will receive the technical training and career mentorship necessary to reach my career goal of becoming a faculty member at an R1 university, while advancing the PGRP goals of studying multiomics responses to abiotic stress, with clear applications to agriculture. Under the mentorship of Dr. Robert VanBuren at Michigan State University, I will develop a holistic toolset of research skills ranging from emerging genomic and molecular techniques to systems and computational biology.

Intellectual merit

The regulatory and spatial dynamics of complex biological processes, such as CCMs, are generally poorly understood. This project will leverage cutting edge single cell and chromatin based approaches to explore how two typically independent CCMs (C_4 and CAM) are regulated within a single leaf. Most of our knowledge of CCMs in eudicots is built upon *Flaveria* for C_4 and *Mesembryanthemum* for facultative CAM, and *Portulaca* will serve as an important model to study the independent evolution of these traits. High resolution surveys from photosynthetic phenotypes to genotypes in *Portulaca* can serve as a roadmap to understand other complex phenomena ranging from abiotic stress tolerance to metabolic network exaptation. These foundational resources and insights will also help establish my independent research career and develop collaborative interactions across the plant science community.

Broader impacts

Adapting agriculture to climate change has been impeded by slow translation of leading edge tools and methods to non-model organisms with phenotypes of interest. Methodological adaptation and development for diverse study systems, such as *Portulaca*, is required to generalize and harness basic research findings. Much effort has been spent engineering C_4 in C_3 crops, such as rice, but in a high CO_2 world with greater variability in precipitation, engineering facultative CAM is equally important. Interactive communication of how basic research meets global challenges inspires future generations of scientists and facilitates the adoption of new technologies. I will engage with persons of all ages on and off campus about how common species can change our understanding of basic principles of biology, such as photosynthesis, and hold keys to agricultural resilience on a changing planet. Michigan State University's Middle School Girl's Math & Science Day affords a unique opportunity to highlight the intersection between coding and botany, and informal events like Lansing's Biology on Tap offer opportunities to discuss C_4+CAM and the benefits of genetic engineering.

"Exploring cell-type regulatory dynamics of CAM and C_4 photosynthesis in *Portulaca*"

1. Introduction

Plant drought resilience is fundamentally related to the exchange of internal water for atmospheric CO_2 required for photosynthesis. C_3 photosynthesis, the ancestral plant photosynthetic pathway, has decreased in efficiency up to 25%[1] due to the reduction in Earth's atmospheric CO_2 over the past 30 million years. This drop in efficiency has spurred the evolution of carbon concentrating mechanisms (CCMs) that augment the concentration of CO_2 around Rubisco relative to the ambient environment[2]. The most common CCMs, C_4 photosynthesis and Crassulacean acid metabolism (CAM), have both evolved through the co-option of deeply conserved anaplerotic reactions[3] and confer advantages beyond meeting the demand for carbon. CO_2 concentration is achieved through a set of carboxylation reactions that generate four carbon acids, which are subsequently decarboxylated in the presence of Rubisco to release CO_2. C_4 is a two-cell photosynthetic system, in which carboxylation takes place in mesophyll cells devoid of Rubisco and decarboxylation occurs in Rubisco-rich bundle sheath cells. In contrast, CAM species conduct gas exchange and carboxylation at night, which lowers transpirational water loss, and decarboxylation occurs during the day while gas exchange is halted. Thereby, C_4 (*spatially*) and CAM (*temporally*) buffer Rubisco from ambient CO_2 levels. High concentrations of CO_2 remove substrate limitations on Rubisco, giving C_4 plants the highest observed photosynthetic rates, and by reducing transpirational water loss, CAM species achieve remarkable water use efficiency. Plants with CCMs also tend to be more nitrogen use efficient and less susceptible to photo-oxidative stress; the benefits of CCMs are evidenced by hyper-productive C_4 crops (e.g., maize, sorghum, sugarcane) and drought resilient CAM crops (e.g., aloe, pineapple)[4,5]. *Portulaca* is one of only four clades reported to use both C_4 and CAM, making it an excellent system to develop a mechanistic understanding of CCM evolution and integration, and ultimately engineer efficient and drought tolerant crops to thrive under dynamic climates.

Although both C_4 and CAM have dozens, if not hundreds, of independent evolutionary origins[6], only two land plant lineages are reported to use both C_4 and CAM (hereafter, C_4+CAM): *Trianthema*[7] and *Portulaca*[8], which both obligately use C_4 photosynthesis and facultatively—and reversibly—exhibit CAM under abiotic stress (e.g., drought, salt, changes in photoperiod). It has been proposed that the rarity of C_4+CAM is in part due to the large overlap in C_4 and CAM metabolites and enzymes[9]; that is, regulation of gene expression and enzyme activity for both C_4 and CAM is assumed to place a large pleiotropic constraint on C_4+CAM evolution. *Portulaca* is an ideal study system because it contains multiple independent origins of C_4+CAM. Two lineages have evolved NADP-type C_4, while a third has evolved NAD-type C_4 and contains C_3-C_4 intermediate taxa. Thus, *Portulaca* provides our only known clade with the diversity in C_4+CAM needed to parse necessary from refining traits, and to understand how different C_4 biochemistries can be integrated with CAM. Since the description of C_4+CAM in *Portulaca*[8], numerous genomic resources have been produced, including a chromosome level genome sequence[10], transformation protocol[11], and transcriptomes from multiple *Portulaca* species[10,12–14]. These research efforts have been successful insofar as they have led to hypotheses of how C_4+CAM may operate, and even suggested potential regulatory mechanisms; however, spatial resolution at the cell-type level is necessary to explicitly test these hypotheses and epigenomic data is needed to understand how the two CCMs are governed in a single leaf.

I will leverage single-cell RNAseq (scRNAseq) and chromatin accessibility (ATACseq) to understand the spatial, temporal, and regulatory dynamics of CCMs in Portulaca. The development of single-cell technologies in plants has lagged behind animals, and only ~20 scRNAseq datasets have been published for a few model plants compared to hundreds of datasets in animals. This discrepancy is largely due to challenges in the preparation of protoplasts or nuclei for single-cell work and the optimizations required for each new species or tissue surveyed. Recent technological improvements and protocols developed in the sponsoring scientist's lab (VanBuren) will enable me to apply single-cell technology to *Portulaca* and gain unprecedented resolution of the processes controlling CCMs in plants. **I will demonstrate how novel multiomics strategies can be adapted to non-model plants, taking C_4+CAM *Portulaca* as a system with need for spatio-temporal resolution at the cell level.**

To identify the distinct genetic elements of each *Portulaca* CCM, I will reconstruct C_4 and CAM gene and regulatory networks using high-resolution scRNAseq and assay for transposase-accessible chromatin using sequencing (ATACseq). Over the course of a CAM-induction experiment, in which water is withheld to induce CAM activity, I will sample leaf tissue for both scRNAseq and ATACseq (Fig. 1A-B). Using cell-type specific markers, such as Rubisco (bundle sheath specific) and PEP carboxylase (mesophyll specific), I will reconstruct the C_4 and CAM gene networks of each cell-type (Fig. 1C). The granularity of scRNAseq will allow me to precisely distinguish C_4 and CAM gene networks across space and time, while ATACseq will reveal corresponding genomic regions with regulatory roles (Fig. 1D). Finally, I will combine gene networks and experimentally verified regulatory motifs from public databases (e.g., JASPAR[15]) that exhibit correlated changes in chromatin accessibility to reconstruct C_4 and CAM gene regulatory networks that encompass genes, transcription factors (TFs), and TF binding sites (Fig. 1E). Finally, by comparing these networks in two species with differing C_4 biochemistries (*P. amilis* and *P. oleracea*), **I will identify the common regulatory networks responsible for CAM in a C_4 plant**, as well as the distinct elaborations of C_4 subtypes. The proposed fellowship will provide training in systems biology and prepare me for a career in biology with rapidly advancing technologies, while furthering the PGRP goals of studying multiomics responses to abiotic stressors–linking plant physiology, ecology, and evolution–with clear applications to agriculture.

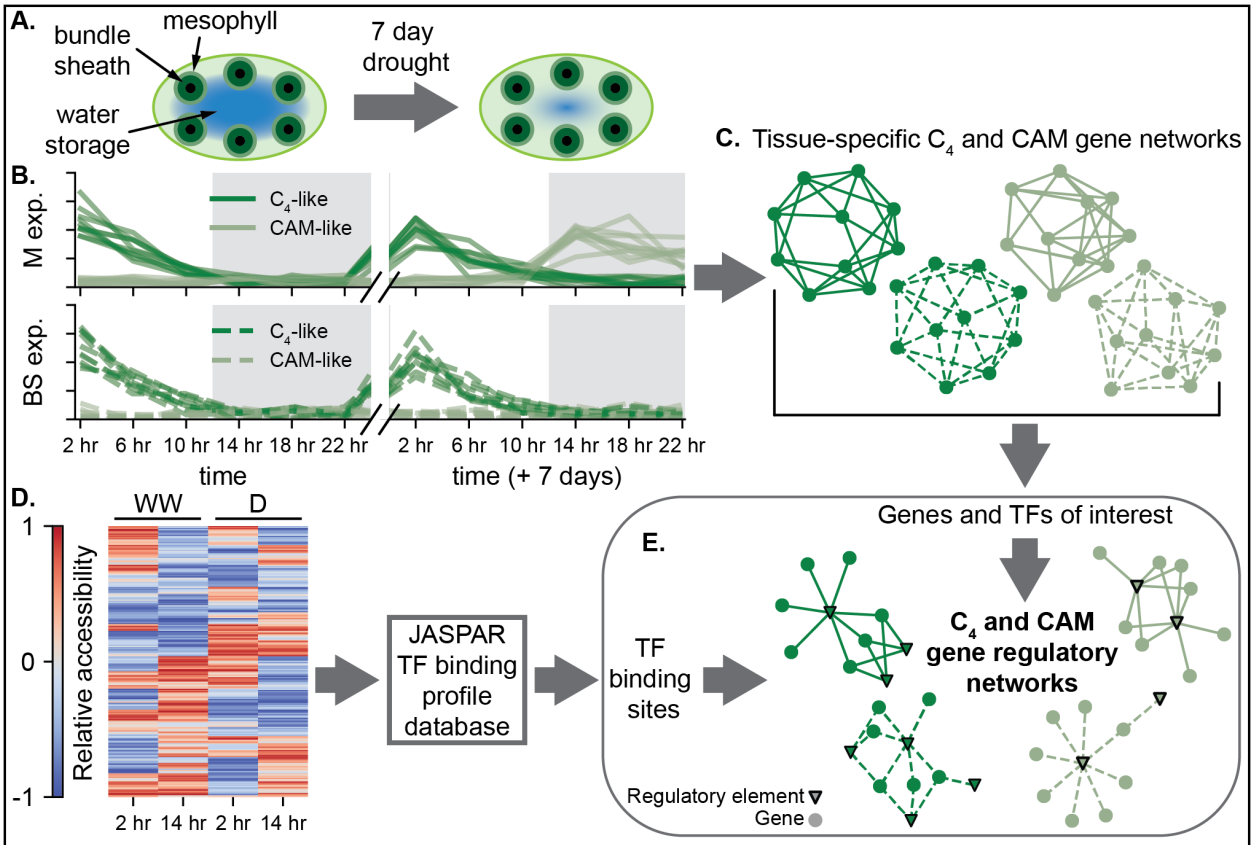

Figure 1. Workflow to identify tissue-specific C_4 and CAM gene regulatory networks. **A**) Cartoon of *Portulaca* leaf cross section showing bundle sheath (BS), mesophyll (M), and water storage tissues over the course of a 7 day drought. **B**) Expected patterns of BS and M gene expression of C_4- and CAM-related genes; their resulting gene correlation networks (**C**) will identify genes and transcription factors (TFs) of interest. **D**) Hypothetical heatmap of relative chromatin accessibility measured by ATACseq at dawn (2 hr) and dusk (14 hr) during well-watered (WW) and drought (D) conditions. Cross-referencing these regions with the JASPAR database of regulatory motifs will reveal putative TF binding sites. Gene regulatory networks will be created by correlating TF binding site states and proximities to C_4 and CAM genes of interest (**E**).

2. Research Statement

Engineering CCMs into crops is considered one of the great challenges of the 21st century[16–19], but fundamental aspects of plant biology, particularly their rigid cell walls, have restricted the omics toolbox common to animal models. This has hindered our study of phenomena like C_4 and CAM, which require precise spatio-temporal resolution. Previous methods to achieve spatial resolution (e.g., immunogold localization[20]) could not delineate the closely related gene family members involved in C_4 and CAM. Furthermore, proteomic studies focus on products that are multiple steps downstream of initial regulation at the genomic level. Previous studies have demonstrated that mesophyll-specific expression of C_4 genes in *Flaveria* is caused by the MEM1 *cis*-regulatory motif[21], and that patterns of CAM gene expression are largely regulated by a relatively small set of circadian clock-related *cis*- and *trans*-factors[19]. However, much less is known about how abiotic stress modulates the magnitude of CAM activity (i.e., how facultative CAM is induced), how C_4 tissue-specific expression is achieved, or if CAM-related genes in *Portulaca* have evolved similar or distinct cell-specific regulatory mechanisms. By combining scRNAseq with ATACseq, I will go beyond identification of tissue-specific C_4 and CAM gene networks to reconstruct the regulatory networks needed for future engineering of CCMs. To this end, I have developed three primary research aims:

I. Measure cell type expression and chromatin dynamics of CCMs in *Portulaca*
II. Identify C_4 and CAM gene expression and regulatory dynamics in *Portulaca*
III. Compare *Portulaca* C_4+CAM systems to identify unique and common features

I will leverage these system-level and evolutionary approaches to address the following questions:

1. How are the distinct C_4 and CAM biochemical pathways governed within a single leaf?
2. What are the minimum elements necessary to induce CAM in a C_4 plant?
3. How different are the networks of distinct C_4+CAM biochemical subtypes?
4. How similar are C_4 and CAM in C_4+CAM *Portulaca* to canonical C_4 and CAM taxa?

Aim I: Measuring cell type expression and chromatin dynamics of CCMs in *Portulaca*

Single-cell and epigenomic methods offer opportunities to describe the roles of individual cell types within organs and how these types are manifested through regulation at various levels. scRNAseq has become an essential tool in delineating cell types and ATACseq is an ideal epigenomic tool to survey regulatory dynamics in non-model organisms because it has a relatively simple protocol, works with low and degraded inputs, and does not require *a priori* knowledge of regulatory targets[22]. These techniques have only recently been adapted from mammalian model organisms to plants and fungi, largely due to difficulties in isolating single cells (protoplasts) with robust cell walls[23]. Hence, most advances in plant single-cell and chromatin accessibility sequencing have relied on isolating nuclei.

Prior to collecting experimental data, I will adapt existing nuclei isolation protocols from the VanBuren Lab that utilize the 10X Genomics Chromium Single Cell platform for *Portulaca*. *Portulaca* are ideal experimental plants, with short life cycles (<6 weeks), self-fertilization and prolific seed production, and small sizes that allow dozens to be grown in a single growth chamber. In brief, pre-cooled labware will be used to chop leaf tissue bathed in Nuclei Isolation Buffer (NIB), which is then incubated in RNase Inhibitor. The homogenate is filtered and centrifuged to produce a nuclei-rich supernatant, which is then captured, resuspended in a wash to remove NIB, and DAPI-stained. Stained nuclei can then be quality-checked using confocal microscopy, and sorted to remove unwanted organelles using fluorescence activated nuclei sorting. For scRNAseq, the nuclei are loaded into the 10X Genomics Chromium controller for library prep and sequencing. Nuclei isolated for ATACseq will instead be subjected to transposition with hyperactive Tn5 for 15 minutes at 37°C to insert sequence adapters into regions of accessible chromatin, and sequenced on a standard short-read platform.

After validating these approaches in one species, *Portulaca amilis*, I will conduct an experiment using two species (*P. amilis* and *P. oleracea*), in which CAM is induced through simulated drought. These two species represent two independent evolutions of C_4+CAM, with distinct C_4 biochemical subtypes, and therefore can be used to identify common and unique facets of C_4+CAM. Leaf tissue will be sampled four

times per day for scRNAseq and twice per day for ATACseq over the course of two days (Fig. 1B): day 0, representing the well-watered C_4-only state, and day 7, representing the drought C_4+CAM state. The associated 10X Genomics CellRanger pipeline includes all steps from scRNAseq read trimming and filtering to cell type classification based on expression patterns. For ATACseq, sequence quality filtering and trimming, alignment to reference genomes, and final quality assessment will be completed with ATACseqQC[24].

Aim II: Identifying C_4 and CAM gene expression and regulatory dynamics in *Portulaca*

Reconstruction of regulatory networks from combined expression and chromatin accessibility is an area of rapid development, but multiple pipelines have demonstrated their inferential power. For example, the Paired Expression and Chromatin Accessibility (PECA) pipeline[25] and the Bayesian network approach AccessTF+TFScore[26] have shown how gene regulatory networks of both *cis*- and *trans*-factors can be infered from scRNAseq and ATACseq. Furthermore, the diffTF pipeline[27] has been used to distinguish between activator and repressor TF roles. To create cell type specific gene regulatory networks, I will first construct gene networks using WGCNA[28] from cell populations classified as bundle sheath and mesophyll (Fig. 1B-C). In the event that scRNAseq is unsuccessful, blunter techniques, such as laser-capture microdissection can be used to isolate and sequence populations of cells. I will assess multiple approaches to cell type classification, including supervised methods (i.e., random forests using marker genes) and unsupervised clustering (e.g., *K*-means, tSNE).

ATACseq data will then be used to identify regulatory motifs, such as TF binding sites, associated with gene networks to ultimately build gene regulatory networks. I will first call ATACseq peaks with HMMRATAC[29] and annotate them using HOMER[30]. Peaks annotated as potential TF binding sites in the JASPAR database[15] will be further interrogated by HINT-ATAC[31] to detect bound TFs through TF footprint analysis (Fig. 1D). The resulting regions will be filtered using differential peak analysis within HOMER to identify TF binding sites associated with each gene network that change in accessibility during the experiment. Finally, using the gene networks and associated TF binding sites, I will establish gene regulatory networks by linking patterns of expression and chromatin accessibility through pipelines such as PECA, AccessTF+TFScore, and diffTF (Fig. 1E).

Aim III: Comparing gene and regulatory networks in *Portulaca*

After replicating the goals of **Aim II** in *P. amilis* (NADP-type C_4+CAM) and *P. oleracea* (NAD-type C_4+CAM), I will establish homology of gene and regulatory elements using OrthoFinder[32] and iqtree[33]. Ancestral gene and genome duplication events have left *Portulaca* with many redundant homologs of core CCM genes[10,12,13]; using metrics such as degree-, bridge-, and eigen-centrality, I will identify which homologs play prominent network roles (e.g., hubs, bridges) in C_4 and CAM in each species. I will also compare the similarity of whole networks by computing largest common subgraphs to find orthologous gene and regulatory interactions. I expect to find moderate variation in membership of *Portulaca* C_4 gene networks because of observed biochemical differences and independent origins. Network comparisons will reveal the degree to which early steps towards C_4 evolution are shared, and I expect higher similarity between gene regulatory networks because temporal and cell-specific expression patterns are common to all *Portulaca*, regardless of conserved, convergent, or parallel histories. And although facultative CAM may have been integrated with C_4 differently in *Portulaca* lineages, it is hypothesized to be ancestral to *Portulaca*, and common CAM and CAM regulatory subgraphs should reveal the minimum requirements to elicit CAM in a C_4 species through abiotic stress, such as drought. I will further compare the C_4 and CAM gene and regulatory networks of *Portulaca* to purely C_4 and CAM species with publicly available RNAseq data, such as *Flaveria* and *Sedum*, respectively, to assess if and how each CCM has been modified to accommodate the other.

3. Research significance and intellectual merit

It was once thought that the photosynthetic system of *Portulaca* was unique, but there have been multiple, recents reports of C_4+CAM from diverse lineages, including grasses[34] and aquatic plants[35,36]. Assaying facultative CAM is difficult because plants' gross morphologies are very similar to C_3 or C_4

plants, and facultative CAM activity does not typically alter commonly sampled isotopic ratios. The extent to which we have underestimated the frequency of C_4+CAM is unclear, but there are numerous C_4 lineages that display some level of succulence that are good candidates. Besides serving as a model for C_4+CAM, *Portulaca* offers insights into both C_4 and CAM separately. Most of our knowledge of C_4 in dicots is built upon *Flaveria*, and similarly *Mesembryanthemum* for facultative CAM. For example, our only studies of bundle sheath and mesophyll gene expression and regulation in dicots come from *Flaveria*[21]. Furthermore, there is still a large knowledge gap in how abiotic stress modulates the amplitude of CAM in facultative CAM species. Diversifying our study systems is essential to distinguish necessary and sufficient traits from lineage-specific refinements. High resolution maps from photosynthetic phenotypes to genotypes in *Portulaca* will serve as models for systems from C_4 and CAM to abiotic stress and gene network exaptation.

4. Broader impacts

The adaptation and development of novel methods for diverse study systems is essential to generalize and harness basic research findings. Plant and fungal research has lagged behind the mammalian research community in multiomics research, despite their essential roles in agriculture, global nutrient cycles, and bioremediation and buffering climate change. Adapting agriculture to climate change is one of the largest challenges facing the international community. Much effort has been spent on engineering C_4 into C_3 crops, such as rice[37,38], but in a high CO_2 world with greater variability in precipitation, engineering facultative CAM may be just as, if not more, important. Breeding or genetic engineering of CAM is only now beginning in C_3 plants, and fine-scale gene and regulatory networks of C_4+CAM in *Portulaca* offer genomic insights and tools for integrating CAM into both C_3 and C_4 crops. In North America, selection on yield in maize has led to decreased drought tolerance[39]—incorporating facultative CAM into crops may greatly improve their water use efficiency and drought resilience.

Interactive communication of how basic research meets the challenges of today inspires future generations of scientists and facilitates the adoption of new technologies. I will advance my research findings beyond the scientific community through outreach programs that the VanBuren Lab is currently engaged in. These programs, including Michigan State University's (MSU) Darwin Discovery Day, MSU's Middle School Girl's Math & Science Day, and Lansing's Biology on Tap bridge the scientific and lay communities across a wide age range on and off campus. *Portulaca* will be used to demonstrate how common, unassuming plants can have surprising traits that change our understanding of tenets of biology, such as plant photosynthesis, and can hold key insights for creating more resilient agricultural systems in the face of climate change. Middle School Girl's Math & Science Day affords a unique opportunity to highlight the intersection between coding and botany, and informal events like Biology on Tap offer opportunities to discuss C_4+CAM and the benefits of genetic engineering.

5. Postdoctoral training and career development

The Postdoctoral Research Fellowship in Biology offers an invaluable opportunity to me as an early career scientist. The increased autonomy as a researcher exploring methods at the forefront of biology while receiving technical and career mentorship will serve as a strong foundation for my career goal of becoming a faculty member at an R1 university. ***Through the PRFB, I will grow a holistic toolset of research skills ranging from horticulture to novel benchtop techniques to multiomics systems biology.*** Working in the VanBuren Lab at MSU's Plant Resilience Institute, I will be trained in high-throughput horticultural methods that involve real-time monitoring of plants' abiotic environments with custom built mini-controllers and laboratory techniques to isolate protoplasts and nuclei, which serve as the entry point to all single-cell molecular methods. I will also be trained in epigenomic methods (i.e., ATACseq), from library preparation through data analysis, and how to incorporate multiomics datasets in a systems biology framework. **More generally, this fellowship will prepare me for a career in biology at a time when method development is far outpacing adoption in non-model organisms; training in the VanBuren Lab will allow me to maintain a flexible, modern, and diverse research program regardless of study system.**

6. Contrasting proposed and dissertation research

The Edwards Lab is guided by the principle of "whole plant evolution"—from molecules to ecosystems—and through my dissertation, which broadly investigated the evolution of CCMs across the plant tree of life, I grew as a comparative biologist. My first dissertation chapter measured the effects of atmospheric CO_2 on rates of CCM evolution over geological timescales in two large flowering plant orders containing tens of thousands of species. The second explored machine learning methods to distinguish C_3, facultative CAM, and obligate CAM species based on simple morphological measurements, in an effort to reduce the phenotyping burden that facultative CAM presents. To facilitate these measurements, I developed computer vision software to segment images and extract cell shapes. Finally, the third chapter focused on expression of C_4 and CAM related genes in *Portulaca amilis* and *P. oleracea* using RNAseq. This chapter provided me with a background in traditional RNAseq analyses, such as differential expression analysis and gene network construction.

The macro-evolutionary perspective of my dissertation work meant treating CCMs as simple, ordered character states projected onto a phylogeny. However, this view belies the highly complex and structured nature of these traits and how they evolve. Through the PRFB, I will learn how to approach CCMs from a systems biology perspective. My training will encompass multiple disciplines I was not able to explore during my dissertation, including horticulture, molecular methods development, and data integration. Importantly, the proposed work will require me to move away from standardized sequencing protocols and analyses to learn how to build methods and tools in order to generate new types of data and novel ways to synthesize them. I believe that a comprehensive understanding of CCMs, from their underlying elements to how these elements evolve in concert, can only be achieved by bridging systems and comparative biology.

7. Justification of sponsoring scientist and institution

The VanBuren lab has expertise in many aspects of photosynthesis and abiotic stress, including drought responses of C_4 plants, desiccation tolerance, and facultative CAM engineering. Furthermore, the VanBuren Lab is pioneering the development of molecular techniques in non-model species, such as single-cell sequencing and ATACseq. As an early career scientist at the intersection of comparative and systems biology, Dr. Robert VanBuren would be an ideal postdoctoral mentor. He has grown a diverse lab at Michigan State University, secured funding for large projects, and mentored students at multiple levels—goals I hope to emulate in my career. The VanBuren Lab is part of MSU's Plant Resilience Institute, which has an expansive network of hundreds of plant science researchers and unrivaled resources for horticultural research; furthermore, the Plant Resilience Institute provides additional support to postdoctoral researchers with external funding. My fellowship proposal fits naturally into the research goals of the VanBuren Lab, as it links multiple ongoing research projects and techniques, and the training and mentorship afforded by Dr. VanBuren will prepare me for a fruitful career as a scientist, mentor, and teacher in evolutionary and systems biology.

8. Fellowship research timeline

Objectives	Year 1				Year 2				Year 3			
Adapt nuclei isolation and sorting protocols to *Portulaca* (Aim 1)												
Validate scRNAseq and ATACseq protocols (Aim 1)												
Conduct CAM induction experiment with 2 *Portulaca* species (Aim 1)												
scRNAseq analysis and gene network construction (Aim 2)												
Chromatin accessibility analyses to identify CCMs TFBSs (Aim 2)												
Construct gene regulatory networks (Aim 2)												
Compare gene and gene regulatory networks of 2 *Portulaca* species (Aim 3)												

References Cited

1. Griffiths, H. Plant biology: designs on Rubisco. 940–941 (2006).
2. Edwards, E. J. & Ogburn, R. M. Angiosperm Responses to a Low-CO_2 World: CAM and C_4 Photosynthesis as Parallel Evolutionary Trajectories. *Int. J. Plant Sci.* **173**, 724–733 (2012).
3. Heyduk, K., Moreno-Villena, J. J., Gilman, I. S., Christin, P.-A. & Edwards, E. J. The genetics of convergent evolution: insights from plant photosynthesis. *Nat. Rev. Genet.* **313**, 1–9 (2019).
4. Kellogg, E. A. C_4 photosynthesis. *Curr. Biol.* **23**, R594–9 (2013).
5. Gilman, I. S. & Edwards, E. J. Crassulacean acid metabolism. *Curr. Biol.* **30**, R57–R62 (2020).
6. Edwards, E. J. Evolutionary trajectories, accessibility and other metaphors: the case of C_4 and CAM photosynthesis. *New Phytol.* **223**, 1742–1755 (2019).
7. Winter, K., Garcia, M., Virgo, A., Ceballos, J. & Holtum, J. A. M. Does the C_4 plant *Trianthema portulacastrum* (Aizoaceae) exhibit weakly expressed crassulacean acid metabolism (CAM)? *Funct. Plant Biol.* **48**, 655–665 (2020).
8. Koch, K. & Kennedy, R. A. Characteristics of Crassulacean Acid Metabolism in the Succulent C_4 Dicot, *Portulaca oleracea* L. *Plant Physiol.* **65**, 193–197 (1980).
9. Sage, R. F. Are crassulacean acid metabolism and C_4 photosynthesis incompatible? *Funct. Plant Biol.* **29**, 775–785 (2002).
10. Gilman, I. S., Moreno-Villena, J. J., Lewis, Z. R., Goolsby, E. W. & Edwards, E. J. Gene co-expression reveals the modularity and integration of C_4 and CAM in *Portulaca*. doi:10.1101/2021.07.07.451465.
11. Ferrari, R. C. *et al.* Developing *Portulaca oleracea* as a model system for functional genomics analysis of C_4/CAM photosynthesis. *Funct. Plant Biol.* **7**, 666–682 (2020).
12. Christin, P.-A. *et al.* Shared origins of a key enzyme during the evolution of C_4 and CAM metabolism. *J. Exp. Bot.* **65**, 3609–3621 (2014).
13. Ferrari, R. C. *et al.* C_4 and crassulacean acid metabolism within a single leaf: deciphering key components behind a rare photosynthetic adaptation. *New Phytol.* (2019) doi:10.1111/nph.16265.
14. Maguvu, T. E., Higuchi, Y., Sugiura, S., Ito, H. & Shibata, M. *De novo* whole transcriptome Analysis of Wingpod Purslane, *Portulaca umbraticola*, and Its Application to the Discovery of Genes Related to Flower Senescence and Opening Rhythm. *Hort. J.* **70**, 97–107 (2020).
15. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–266 (2018).
16. Borland, A. M. *et al.* Engineering crassulacean acid metabolism to improve water-use efficiency. *Trends Plant Sci.* **19**, 327–338 (2014).
17. Yang, X. *et al.* A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytol.* **207**, 491–504 (2015).
18. Schuler, M. L., Mantegazza, O. & Weber, A. P. M. Engineering C_4 photosynthesis into C_3 chassis in the synthetic biology age. *Plant J.* **87**, 51–65 (2016).
19. Schiller, K. & Brautigam, A. Engineering of Crassulacean Acid Metabolism. *Annu. Rev. Plant Biol.* (2021).
20. Guralnick, L. J., Edwards, G., Ku, M. S. B., Hockema, B. & Franceschi, V. R. Photosynthetic and anatomical characteristics in the C_4-crassulacean acid metabolism-cycling plant, *Portulaca* grandiflora. *Funct. Plant Biol.* **29**, 763–773 (2002).
21. Gowik, U. *et al. cis*-Regulatory elements for mesophyll-specific gene expression in the C_4 plant *Flaveria trinervia*, the promoter of the C_4 phosphoenolpyruvate carboxylase gene. *Plant Cell* **16**, 1077–1090 (2004).
22. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
23. Cole, B. *et al.* Plant single-cell solutions for energy and the environment. *Commun. Biol.* **4**, 962 (2021).
24. Ou, J. *et al.* ATACseqQC: a Bioconductor package for post-alignment quality assessment of

ATAC-seq data. *BMC Genomics* **19**, 169 (2018).
25. Duren, Z., Chen, X., Jiang, R., Wang, Y. & Wong, W. H. Modeling gene regulation from paired expression and chromatin accessibility data. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4914–E4923 (2017).
26. Ji, Z. *et al.* Genome-scale identification of transcription factors that mediate an inflammatory network during breast cellular transformation. *Nat. Commun.* **9**, 2068 (2018).
27. Berest, I. *et al.* Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF. *Cell Rep.* **29**, 3147–3159.e12 (2019).
28. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
29. Tarbell, E. D. & Liu, T. HMMRATAC: a Hidden Markov ModeleR for ATAC-seq. *Nucleic Acids Res.* **47**, e91 (2019).
30. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
31. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
32. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
33. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
34. Ho, C.-L., Chiang, J.-M., Lin, T.-C. & Martin, C. E. First report of C_4/CAM-cycling photosynthetic pathway in a succulent grass, *Spinifex littoreus* (Brum. f.) Merr., in coastal regions of Taiwan. *Flora* **254**, 194–202 (2019).
35. Zhang, Y. *et al.* Biochemical and biophysical CO_2 concentrating mechanisms in two species of freshwater macrophyte within the genus *Ottelia* (Hydrocharitaceae). *Photosynth. Res.* **121**, 285–297 (2013).
36. Li, P. *et al.* Bicarbonate-use by aquatic macrophytes allows a reduction in photorespiration at low CO_2 concentrations. *Environ. Exp. Bot.* **188**, 104520 (2021).
37. Sheehy, J. E., Mitchell, P. L. & Hardy, B. Charting New Pathways to C_4 Rice. (2008) doi:10.1142/6560.
38. Wang, S., Tholen, D. & Zhu, X.-G. C_4 photosynthesis in C_3 rice: a theoretical analysis of biochemical and anatomical factors. *Plant Cell Environ.* **40**, 80–94 (2016).
39. Lobell, D. B. *et al.* Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science* **344**, 516–519 (2014).

NSF BIOGRAPHICAL SKETCH

NAME: Gilman, Ian

NSF ID: 000867730@nsf.gov
ORCID: 0000-0002-0390-9370

POSITION TITLE & INSTITUTION: Fellow, Yale Institute for Biospheric Studies

(a) PROFESSIONAL PREPARATION -(see PAPPG Chapter II.C.2.f.(a))

INSTITUTION	LOCATION	MAJOR / AREA OF STUDY	DEGREE (if applicable)	YEAR YYYY
Bucknell University	Lewisburg, PA	Physics	BS	2015
University of Idaho	Moscow, ID	Biological Sciences	MS	2017
Yale University	New Haven, CT	Ecology & Evolutionary Biology	PHD	2022

(b) APPOINTMENTS -(see PAPPG Chapter II.C.2.f.(b))

2020 - 2021 Fellow, Yale Institute for Biospheric Studies, New Haven, CT

2017 - 2018 Henry Daggett Hooker Memorial Fellow, Yale Graduate School of Arts and Science, New Haven, CT

(c) PRODUCTS -(see PAPPG Chapter II.C.2.f.(c))

Products Most Closely Related to the Proposed Project

1. Gilman IS, Moreno-Villena JJ, Lewis ZR, Goolsby EW, Edwards EJ. Gene co-expression reveals the modularity and integration of C_4 and CAM in *Portulaca*. Plant Physiology. Forthcoming; (in revision).
2. Moreno-Villena JJ, Zhou H, Gilman IS, Tausta S, Cheung C, Edwards EJ. Spatial resolution of an integrated C_4+CAM photosynthetic metabolism. bioRxiv [Preprint]. 2021 November 26.
3. Gilman IS, Goolsby EW, Edwards EJ. An annotated chromosome-level genome of *Portulaca amilis*. [Internet]. Phytozome: Joint Genome Institute; 2021 August. Available from: https://phytozome-next.jgi.doe.gov/info/Pamilis_v1_0
4. Heyduk K, Moreno-Villena JJ, Gilman IS, Christin PA, Edwards EJ. The genetics of convergent evolution: insights from plant photosynthesis. Nat Rev Genet. 2019 Aug;20(8):485-493. PubMed PMID: 30886351.
5. Gilman IS, Edwards EJ. Crassulacean acid metabolism. Curr Biol. 2020 Jan 20;30(2):R57-R62. PubMed PMID: 31962074.

Other Significant Products, Whether or Not Related to the Proposed Project

1. Gilman IS, Edwards EJ. MiniContourFinder: lightweight segmentation software for biological images. Bioinformatics. Forthcoming; (in review). Available from: https://github.com/isgilman/MiniContourFinder
2. Lee AK, Gilman IS, Srivastav M, Lerner AD, Donoghue MJ, Clement WL. Reconstructing Dipsacales phylogeny using Angiosperms353: issues and insights. Am J Bot. 2021 Jul;108(7):1122-1142. PubMed Central PMCID: PMC8362060.
3. Martine CT, Jordon-Thaden IE, McDonnell AJ, Cantley JT, Hayes DS, Roche MD, Frawley ES, Gilman IS, Tank DC. Phylogeny of the Australian Solanum dioicum group using seven nuclear genes, with consideration of Symon's fruit and seed dispersal hypotheses. PLoS One.

2019;14(4):e0207564. PubMed Central PMCID: PMC6472733.

4. Gilman IS, Tank DC. Species Tree Estimation using ddRADseq Data from Historical Specimens Confirms the Monophyly of Highly Disjunct Species of *Chloropyron* (Orobanchaceae). Systematic Botany. 2018 September 10; 43(3):701-708. DOI: 10.1600/036364418X697418

(d) SYNERGISTIC ACTIVITIES -(see PAPPG Chapter II.C.2.f.(d))

1. 2019-2021 Member of the Botanical Society of America's Investment Committee
2. 2016-2019 Mentor at Planting Science (plantingscience.org) to K-12 students as they created and conducted plant-based experiments
3. 2016 Mentor for the Botanical Society of America's Preparing Leaders and Nurturing Tomorrow's Scientists (PLANTS) program for promoting diversity in botany
4. *Ad hoc* reviewer for *Molecular Ecology*, *Planta*, *PLoS ONE*, and *Nordic Journal of Botany*

NSF CURRENT AND PENDING SUPPORT

PI/co-PI/Senior Personnel: Gilman, Ian

PROJECT/PROPOSAL PENDING SUPPORT

1. Project/Proposal Title: Cell-level dynamics of a rare photosynthetic pathway: regulation of CAM in C4 Portulaca (THIS PROPOSAL)

   Proposal/Award Number (if available):

   Source of Support: NSF

   Primary Place of Performance: Michigan State University

   Project/Proposal Support Start Date (if available): 2022/07

   Project/Proposal Support End Date (if available): 2025/07

   Total Award Amount (including Indirect Costs): $207,000

   Person-Month(s) (or Partial Person-Months) Per Year Committed to the Project:

Year	Person-months per year committed
2023	12
2024	12
2025	12

   Overall Objectives: 1. Measure cell type expression and chromatin dynamics of CCMs in Portulaca; 2. Identify C4 and CAM gene expression and regulatory dynamics in Portulaca; 3. Compare Portulaca C4+CAM systems to identify unique and common features

   Statement of Potential Overlap: None

Data Management Plan

Collection, storage, and maintenance of all data relevant to this proposal will follow the FAIR principles of Findable, Accessible, Interoperable, and Reusable. Data (including metadata) will generally be made accessible upon publication of preprints, and made findable, interoperable, and reusable through appropriate host databases and documentation. The proposed work will generate large-scale genomic datasets, analytical software, and publications. Below are the data management and dissemination plans for each aim:

Aim	Anticipated outcomes	Data dissemination
Measure cell type expression and chromatin dynamics of CCMs in *Portulaca*	scRNAseq and ATACseq datasets	All raw RNAseq data will be deposited on NCBI's SRA, and processed datasets will be archived on DataDryad and on the MSU HPCC
Identify C_4 and CAM gene expression and regulatory dynamics in *Portuaca*	Transformed and integrated multi-scale datasets, open chromatin peaks, gene co-expression networks, putative *cis*-elements	Processed datasets will be archived on DataDryad, code will be hosted on GitHub, results will be disseminated through peer reviewed publication
Comparing gene and regulatory networks in *Portulaca*	Integrated co-expression networks	All metadata and processed datasets will be archived on DataDryad. Code will be hosted on GitHub and or Jupyter notebooks, results will be disseminated through peer reviewed publication

Data Storage

All sequence data (scRNAseq and ATACseq) will be deposited publicly to the NCBI sequence read archive (SRA) in BioProjects with detailed descriptions of the sources and experimental designs. scRNAseq reads and experimental metadata will also be deposited on EMBL's Single-Cell Expression Database (ebi.ac.uk/gxa/sc). Raw reads will be used to generate reference products, including cell-type specific transcriptomes and atlases of transcription factor binding motifs. Expression profiles from the scRNAseq experiment will be uploaded to PLEXdb and gene co-expression networks and analyzed genome-wide datasets will be deposited on CyVerse for long-term storage. Sequence data will be deposited immediately after generation with a "hold until publication" embargo. Data will be released prior to publication by request with strict adherence to the Toronto agreement that states that no whole-genome analyses can be performed until the initial results are published. All raw data, intermediate files, and final outputs will also be stored long-term at the High Performance Computer Cluster (HPCC) at MSU in addition to public deposition.

Data Sharing

Seeds collected from experimental plants will be kept in cold storage at 5°C and be made available to the research community upon request. Analytical software and code from this project will be hosted on the public repository GitHub under a dedicated channel for this project. Step-by-step walkthroughs from raw read demultiplexing to reference assembly and downstream analyses will be hosted on the project GitHub using repository Wikis. This repository will also hold any custom code and tools for data wrangling, analysis, and visualization that are developed throughout the fellowship. All GitHub repositories will be protected under a GNU General Public License that guarantees end users are free to use, share, and modify their contents.

Chapter 1: The phylogenetic distribution of carbon concentrating mechanisms

Measurements of C_4 diversity across the plant tree of life have yielded insights into the most common and variable aspects of C_4 evolution, the order of trait assembly, and relationships between C_4 and climate. Analogous studies of CAM, such as traits that enable CAM evolution and the number of CAM origins, have been impeded by the diversity of CAM lineages and cryptic nature of facultative CAM (hereafter C_3+CAM). The recent surge in research that employs real-time physiological measurements (e.g., gas exchange, nocturnal acid accumulation, mRNA expression) has greatly increased the counts of CAM lineages and more precisely located transitions to CAM on phylogenies. In this chapter I create a comprehensive database of plant photosynthetic pathways across vascular plants and use this information to measure transition rates from C_3 to C_4 and from C_3 to CAM in two large angiosperm clades, the Poales and Caryophyllales. By time-calibrating these phylogenies, I assess the relationships between C_4 and CAM evolution and historic levels of CO_2 over the past 100 MY.

Chapter 2: Delimitation of photosynthetic types using leaf ultrastructure and machine learning

Morphological innovations play a central role in C_4 and CAM, but shifts in morphology during the evolution of CAM are poorly understood with respect to their magnitude, timing, and, to some degree, necessity. Morphology has been described as the "rate-limiting step" in carbon concentrating mechanism (CCM) evolution, with morphological shifts hypothesized to occur early and late in C_4 and CAM evolution, respectively. Obligate CAM species tend to show increased succulence that manifests as thicker leaves, expanded water storage tissues, and decreased intercellular airspace (IAS). However, it is unclear to what degree C_3+CAM plants are morphologically distinguishable from their C_3 or CAM relatives, if they are at all. The broad category of C_3+CAM contains plants that constitutively perform low amounts of CAM relative to C_3 photosynthesis and plants that exhibit CAM solely in response to stress; therefore the magnitude and pattern of CAM expression can vary greatly, as do their morphologies. For example, many C_3+CAM orchids and members of the Crassulaceae have gross morphologies similar to their obligate CAM relatives, but other facultative CAM species, such as some bromeliads (e.g., *Guzmania*), appear morphologically indistinguishable from C_3 species. In this chapter I develop image segmentation software to facilitate morphological measurements from histological slides. Using this software, I construct a novel morphological dataset of photosynthetic tissue from the cacti and their close relatives, and combine it with publically available datasets from orchids and bromeliads to test for significant differences in C_3, C_3+CAM, and obligate CAM morphologies. Using the machine learning methods of random forests and support vector machines, I identify a set of morphological measurements predictive of photosynthetic phenotype.

Chapter 3: The *Portulaca* genome and transcriptomics of facultative CAM in a C_4 plant

Although both C_4 and CAM have dozens, if not hundreds, of independent evolutionary origins, only four plant lineages have been reported to use both C_4 and CAM (hereafter, C_4+CAM): *Trianthema*, *Ottelia*, *Spinifex*, and *Portulaca*. *Portulaca* obligately uses C_4 photosynthesis and facultatively—and reversibly—exhibits CAM under abiotic stress (e.g., drought, salt, changes in photoperiod). It has been hypothesized that the rarity of C_4+CAM is in part due to the large overlap in C_4 and CAM metabolites and enzymes; that is, regulation of gene expression and enzyme activity for both C_4 and CAM is assumed to place a large pleiotropic constraint on C_4+CAM evolution. In this chapter I explore this hypothesis by measuring changes in photosynthetic gene expression in two *Portulaca* species (*P. amilis* and *P. oleracea*) during CAM-induction. I annotate a genome assembly for *P. amilis*, which addresses fundamental questions related to the number of CCM-related gene copies and guides transcriptome assembly. ***This work provided limited insights into the spatial and temporal dynamics of CCMs in Portulaca, and my proposed fellowship project would leverage advanced single-cell and chromatin dynamic datasets to provide a deeper understanding of this complex phenotype. The fellowship will also allow me to develop a new set of skills in molecular biology, systems biology, and computational biology.***

Sponsoring Scientist Statement

Re: NSF Postdoctoral Research Fellowships in Biology (PRFB)

To: Post-Doctoral Fellowship Evaluation Panel

Overview: The VanBuren lab is highly active in the CAM photosynthesis and drought stress research communities and offers a diverse set of expertise in functional and comparative genomics, ecophysiology, biochemistry, computational biology, and molecular biology. The breadth of expertise, state of the art facilities, and targeted mentoring plan will provide a nurturing training environment to prepare Dr. Ian Gilman for his career goal of developing his own research program. Dr. Robert VanBuren will serve as the primary mentor of Dr. Gilman and Michigan State University will serve as the host institution. MSU has a large and diverse plant science community with 170+ faculty spanning all aspects of basic and applied research in both crop and model species. The MSU campus has state-of-the-art genomics, computational, microscopy, mass spectrometry, tissue culturing and greenhouse/growth facilities that will be available for use in this project. MSU has a large and diverse postdoctoral researcher community with over 200 postdocs working in plant science labs. This will provide an ideal environment for Dr. Gilman to develop the research, educational, communication and mentoring skills he will need for the next stage of his career.

Research in the VanBuren Lab: The VanBuren lab is broadly interested in the evolution of drought tolerance mechanisms in plants. Work in the lab centers on the evolution of CAM photosynthesis and desiccation tolerance as well as drought stress mechanisms in C4 cereals and their wild relatives. For CAM photosynthesis related projects, we apply an integrated comparative genomic, quantitative genetic, physiological, and evolutionary approach to understand the genetic elements underlying CAM. The VanBuren lab uses pineapple as a model to study the mechanisms controlling constitutive CAM and *Sedum album* to study the evolution of facultative CAM photosynthesis. The VanBuren lab has recently developed protocols for single cell and single nuclei RNAseq from leaf tissues of diverse C4 and CAM species. This includes an optimized protoplast isolation protocol for recalcitrant tissue and a general nuclei isolation protocol. Single cell libraries have been constructed using both the Illumina 10x Genomics Chromium platform and a custom built DropSeq system in the VanBuren Lab. We have also identified cell type specific markers that can distinguish several important cell types in leaves including pavement cells, guard cells, mesophyll, bundle sheath, and vasculature related cells that are useful in C4 grasses (maize and sorghum) and pineapple. ***We have the personnel, equipment, and computational expertise to conduct the single cell RNAseq experiments proposed by Dr. Ian Gilman.*** This project is currently supported by startup funding from MSU and has no overlap with the project proposed by Dr. Gilman.

Research on desiccation tolerance in the VanBuren lab is supported by the National Science Foundation (NSF-MCB: 1817347). This project is focused on the origin of desiccation tolerance in C4 grasses and has no overlap with the work proposed by Dr. Gilman. Under this project, we have developed protocols for ATACseq, ChIPseq, and other chromatin-based techniques for several grasses in different tissues. ***We have the computational and molecular expertise to collect the ATACseq data proposed by Dr. Ian Gilman***. Research on stress tolerance in C4 cereals is supported by funding from the USDA-NIFA (2022-67013-36118). Other major projects in the lab included natural variation of drought tolerance in sorghum and the orphan grain crop Tef. Phenotyping, physiology, and quantitative genetics expertise for these projects in the VanBuren lab will aid Dr. Gilman in designing and analyzing the experiments in his project and help expose him to new areas including applied and field-based research. The genomic resources and computational tools developed in the VanBuren lab will be available to Dr. Gilman and will help him establish various *Portulaca species* as models for carbon concentration mechanisms.

Computational resources are available through the MSU High Performance Computing Center (HPCC). The HPCC has 600 available nodes with 65,208 total cores ranging from 128 Gb to 2 Tb of memory per node. The VanBuren lab has 80 Tb of dedicated storage space and priority access to two 128 processor machines with 1 Tb of memory through the MSU HPCC. This should be more than sufficient for any computational needs of Dr. Gilman.

Training program for Dr. Ian Gilman: Dr. VanBuren will work closely with Dr. Gilman to create an individualized development plan (IDP) to prepare Dr. Gilman for establishing his own independent research program. Dr. Gilman's skills will be assessed using the IDP tool from The American Association for the Advancement of Science (AAAS). This will help identify strengths and areas to focus on for improvement. Mentoring goals within the IDP will be sent to the MSU Office for Postdoctoral Training and Professional Development Opportunities to ensure they are consistent with the University guidelines. Specific activities will be incorporated with the mentoring plan outlined below.

Research training: The research outlined by Dr. Gilman leverages expertise in the VanBuren lab and the broader plant science community at MSU. This includes strengths in genomics, physiology, molecular biology, and computational biology in the VanBuren lab. Dr. Gilman brings his own unique expertise in phylogenetics, evolutionary biology, and genetics, which should enrich the host lab. Within the VanBuren lab, Dr. Gilman will receive training on single cell gene expression, epigenetics, systems biology, data science, and computational biology. Dr. VanBuren has an "open door" policy and he meets with all lab members weekly. Dr. Gilman will receive similar mentoring and be encouraged to pursue additional training opportunities such as short-courses offered through Cold Spring-Harbor or annual meetings [e.g. American Society of Plant Biologists (ASPB), Plant and Animal Genome (PAG)]. Other scientists at MSU who can provide relevant supporting expertise include Dr. Shin-Han Shiu (computational biology), Dr. Berkely Walker (physiology and biochemistry), Dr. Kevin Childs (bioinformatics), Dr. David Lowry (natural diversity of stress responses), Dr. Addie Thomson (drought tolerance and quantitative genetics), and Dr. Patrick Edger (comparative genomics). Dr. VanBuren has active collaborations with most of these faculty members and participates in a joint lab meeting on computational plant biology. This collaborative network and joint lab meetings provide a forum for troubleshooting issues related to bioinformatics and computational biology. Dr. VanBuren is actively involved in an NSF-NRT training grant (#1828149) for cross-disciplinary mentoring in plant and computational sciences. Dr. Gilman will have the opportunity to participate in training modules, reading groups, and outreach actives associated with this exciting program. Dr. Gilman will be encouraged to disseminate his research findings through oral and poster presentations as well as publication in peer reviewed journals with an emphasis on first authored papers. The travel stipend for the NSF fellowship will allow for yearly travel to present and network at major conferences. Dr. Gilman will work closely with Dr. VanBuren during the drafting, submission, and revision of manuscripts.

Teaching and outreach training: Dr. Gilman will have the opportunity to conceptualize, design, and develop outreach activities at MSU. Ongoing outreach activities within the VanBuren lab include teaching modules within the MSU 4-H Children's garden and horticultural garden which together attract ~130,000 visitors annually of all age groups. The VanBuren lab is also involved in activities for Darwin Day and a Raspberry Pi Jam as part of the NSF-NRT training grant in computational plant biology. Each of these outreach activities have dedicated staff for assessment and evaluation of learning outcomes which will provide Dr. Gilman with useful feedback for improvement. Dr. Gilman will have opportunities to further develop his teaching skills through developing lectures and formal training programs. At MSU, Dr. Gilman will be encouraged to take the Pathways to Scientific Teaching course by the distinguished plant science educator Dr. Diane Ebert-May.

Career development training: The MSU Postdoctoral Association hosts a professional development workshop series and provides individualized mentoring plans through the Center for Academic and Future Faculty Excellence. The Postdoctoral Association provides opportunities to nurture vocational aspects of post-doctoral development through participation in: scientific seminar organization, professional networking and speaking, manuscript and proposal preparation, ethical training and mentoring of others. Dr. Gilman is encouraged to attend development workshops at conferences in addition to these internal opportunities. Dr. VanBuren will also provide mentoring during future job searches including reviewing application materials and practice talks.

After completion of this fellowship, Dr. Gilman is encouraged to pursue independent research on CCMs in Portulaca, and he can take any element of the project with him when he establishes an independent research program. The genomic, and systems-level resources he is developing could form the basis of a competitive independent NSF-IOS or NSF-MCB proposal.

Best regards,

Dr. Robert VanBuren
Assistant Professor
Plant Resilience Institute
Department of Horticulture
Michigan State University