Clarifying Connections: Seeking to understand ways Science, society, plants, and microorganisms intertwine

by

Lorenzo Washington

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Henrik Scheller, Chair

Professor Devin Coleman Derr

Professor Mary Firestone

Summer 2024

Clarifying Connections: Seeking to understand ways Science, society, plants, and microorganisms intertwine

By Lorenzo Washington

Abstract

Clarifying Connections: Seeking to understand ways Science, society, plants, and microorganisms intertwine

by

Lorenzo J. Washington

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Henrik V. Scheller, Chair

One of the largest and longest standing societal challenges is acquiring enough food to support the human population. Our current era has the added complexities of the population nearing 9 billion, a globalized industrial agricultural production system, and a clearer understanding of how environmentally harmful and unsustainable that system is. While these issues are multifaceted, agriculture is fundamentally coupled with plant and microbial biology and there has been a diversity of avenues pursued by researchers seeking potential solutions and disruptive innovations to address our need to provide healthy food for the world without destroying it in the process. Modern agricultural practices currently require high energetic costs through extensive irrigation, application of synthetic fertilizers and pesticides, and the use of large, mechanized equipment. The lack of biodiversity in large-scale monoculture systems also presents issues in crop nutritional quality, disease pressure, and soil health. Recent advancements in our understanding of plant root systems and the relationships they form with soil microbiota have demonstrated a wealth of potential applications regarding these issues as well as in further illuminating the complex processes at play in plant health and development in an ecological context. It has become increasingly clear that plant roots are essential components to local and global nutrient cycling and drive much of the microbial activity in soils. To ensure a future with sustainable, climate-resilient agriculture which enables nutritious diets, we must continue to explore the wealth of knowledge underground. This dissertation seeks to clarify some of the socioeconomic reasons why we grow food the way we do and the limitations of material available for study, as well as underlying biochemical and genetic mechanisms which influence plant root biology and their relationships with microorganisms. The results of this research contribute to the growing body of knowledge in root biology and a paradigm shift in how we understand plants and agriculture to be connected to the wider ecosystem.

Chapter 1 seeks to understand the historical and contemporary contexts which influence the scale, scope, and direction of research in agricultural and plant sciences. Analyzing the socioeconomic institution of modern Science, from the conditions leading

to the Scientific Revolution in the 16th century to modern day, reveals the conserved influence of western European colonial-imperialism on global agricultural production. From its inception, the institution of Science is shown to be integral to the expansion and maintenance of Western imperial powers materially and socially, by driving critical revenue generation via the enabling and adoption of cash crop agriculture and through controlling the value and direction of intellectual pursuits. From the 19th century we see Science increasingly entangled with emerging capitalist corporate and state interests, which further entrench practices that began with cash crop agriculture and prove to be detrimental to environmental health while distancing crop production from nutritional needs. I describe how these relationships ultimately limit the resources we have available for research in the modern day, hindering our abilities to address the myriad socio-scientific challenges of this century. This clarity is important when making the critical and strategic assessments necessary to direct on-going and future research and teaching efforts.

Chapter 2 is an effort to detail that plant roots are essential to plant health and adaptation as well as important contributors to numerous ecological and biogeochemical processes. Despite this, they have been comparatively understudied aspects of plant biology. Recent developments have indicated that a better understanding of root systems can reveal breeding and engineering applications towards plants more well-suited for low-input agricultural systems and frequent stresses. I highlight four co-authored publications concerning the study of root systems: Detailing the limitations and considerations of root system study as well as demonstrations of progress to address these challenges, metabarcoding combined with genomics and data analytics programs to inform the basis of rhizosphere microbiome heritability in *Sorghum bicolor*, constitutive promoters to fine tune gene expression in synthetic biology with demonstrated function in root systems, and utilizing root system imaging methods and software to inform the effect of a bacterial metabolite on plant-bacterial associations in *Arabidopsis thaliana*. These studies hope to provide examples of and inspire further efforts to understand root systems, thus enabling improvements in plant performance under new and changing agricultural practices better suited for the world we live in today.

Chapter 3 seeks to further illuminate the underlying mechanisms of pectin biosynthesis. Pectin is an abundant component of plant cell walls demonstrated to be important in cell to cell adhesion, plant development, and a variety of signaling pathways. Understanding has so far remained limited due to the wide diversity and redundancy of enzymes involved in the process. The pectin component rhamnogalacturonan I (RG-I) is biosynthesized in part by rhamnosyltransferases (RRTs) in the GT106 family. While it is known that RRTs are expressed in a range of tissues, few studies have demonstrated their function outside of RG-I biosynthesis in seed coat mucilage. In this study we show *Lotus japonicus* RRT1 contributes to the addition of rhamnose monosaccharide residues to RG-I in root tissues. Mutants contained ~19% less rhamnose in their roots and pectin from aboveground tissues had less galactose and more xylose. RG-I in root tissue from *Ljrrt1* also had a larger molecular weight and altered structure compared to wild type (WT) Gifu plants, but this was not due to transcriptional differences in other GTs responsible for RG-I biosynthesis. Mutants exhibited altered root morphology,

impacted stem and root growth, and impairment of nodule formation when inoculated with *Mesorhizobium loti*. These findings constitute the first demonstration of RRT function in vascular plants outside of seed coat mucilage and contribute to the increasingly nuanced understanding of RG-I in cell wall biosynthesis and intersecting processes.

Chapter 4 is an effort to characterize an unknown subclass of plant GTPase-related signaling proteins which appear to influence symbiotic relationships formed in root systems. Plants possess a unique class of heterotrimeric Gα subunits called extra-large GTPases (XLGs) which contribute to numerous developmental and stress responses. XLGs have an uncharacterized N-terminal domain, a Gα-like C-terminal domain, and overlapping and distinct functions compared to conventional Gα subunits. In this study, we identified homologs of XLG3 in *Lotus japonicus* responsive to rhizobial and mycorrhizal symbiosis. However, these proteins were approximately one-third the size of conventional XLGs and only aligned to the N-terminal domain, containing a putative NLS and the cysteine-rich domain of unknown function. Multiple sequence alignment and phylogenetic analysis determined SXLGs did not share domains with other mono- or heterotrimeric G-protein classes and exhibited a pattern of duplication and neofunctionalization typical of genes involved in symbiotic signaling pathways. Transient expression of *LjSXLG*s in tobacco demonstrated their potential for localization to the plasma membrane, nucleus, and nucleolus. Analysis of *L. japonicus sxlg2* mutants revealed transient impairment of immature nodule formation in a destructive experimental setup and inhibition of infection events in a nutrient-limited non-destructive experimental setup, with no observed difference in nodule maturation rate. Additionally, *sxlg2* mutants showed a potential impairment of the root growth response in N-limited conditions. SXLGs present an ideal opportunity to better understand the evolution, function, and structure of XLGs and are another example of G-protein involvement in symbiotic relationships.

Ultimately, this research supports growing efforts to develop more resilient and sustainable agricultural practices through a focus on root systems biology by providing assessments of the technical and methodological resources in the field, demonstrating dynamics of pectin biosynthesis in root tissue, and uncovering new elements of plant symbiotic and G-protein related signaling pathways. These findings will promote further mechanistic and evolutionary discoveries in aspects of root biology that remain filled with questions and unknowns, but also notable potential in the development of roots more amenable to regenerative agricultural approaches, maximizing benefits from microbial associations, and utilization in the biosynthesis of valuable products and biofuels. The situating of historical and contemporary socioeconomic contexts that have heavily influenced the progression of agriculture and plant sciences over the past half a millennium is a critical addition to our ability to wholly assess the materials, practices, and technologies available for further research. This clarity is essential if we truly wish to address the systemic challenges of our era. Technical solutions have limited ability to resolve complex socio-scientific issues without understanding the broader social contexts they were born from and operate in. In the same way that systems biology has begun to permeate many scientific disciplines, our perspectives must shift to accommodate the nuance and complexity of how interconnected this world is

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, this dissertation is dedicated to my parents, Leandra and Rodney Washington, who gave me everything I needed to grow into the person I am today. Y'all planted seeds for me that continue to bear fruit and provide shade to this day and have always been there when I needed you, whether I knew it or not.

Throughout my studies I've been blessed with an abundance of love and support from many people in many ways. This hardly captures the extent of the village which enabled me to do things as I have done, and I hope everyone who deserves their flowers from me will come to know it one way or another.

To Drs. Henrik Scheller and Norval Hickman, thank you for the years of advice, guidance, mentorship, and other forms of support. Henrik, as my advisor you have ensured my access to a phenomenal wealth of resources and experiences which have helped me grow immensely as a researcher. You've been patient with me when I've let my focus drift beyond the bench and have consistently provided invaluable insight towards my projects, thank you. Norval, as a mentor you helped me navigate the difficulties of being Black in academia and were a constant reminder that, even in the hardest moments, I would see this through to the end. Thank you for everything.

To my thesis committee, my partner Rachall, and the many friends and colleagues that have been sources of inspiration, courage, relief, knowledge, advice, fun, and then some thank you so much. Without y'all I would never have developed my analysis, politics, and sense of self to the extent I have or had the support to pursue my scholarly interests as I did. I have learned and grown so much thanks to y'all and have no way to properly put it into words. Naming all of you and what you've done for me would be its own dissertation, but y'all know who you are and what we been doing together. I look forward to making many more memories and realizations in community with y'all.

To my family: my sister Zada, my grandparents, aunties, uncles, cousins, and the friends who have seen every version of me. Y'all are the garden I've been rooted in since Day One, thank you for believing in me even when I didn't, making sure I knew the importance of what I was accomplishing, and being my Home to come back to.

The Black and Queer communities I've been privileged to be part of here in the Bay have held me and shown me what is possible if we let ourselves think beyond the cages this world has put around us. We stayed true to ourselves in all we did and made sure to throw a lotta joy in there too; it helped me grow in ways I didn't know I needed. I came to Berkeley not knowing what a transformation I would undergo thanks to you both and because of y'all I'd do it all over again. The little moments of acknowledgment and endless love, care, and support have meant so much to me.

An immense thank you to all the staff that have enabled my work and made sure I felt welcomed in these spaces. Sadiki Showers, Mary Agnitsch, Joanne Straley, and Rocio Sanchez were especially important beyond their logistical support.

Chapter 1. The socioeconomic institution of Science has limited the material resources available for agricultural research

1.0 Chapter Preface

This text is an effort to collect thoughts and analysis developed over the course of my doctoral education regarding the larger societal structures which we — scientists, academics, educators, students, etc. — operate within and are influenced by as we carry out our research and teaching. It was born through many readings, discussions, and personal experiences I had alone as well as in community with a wide array of people occupying different positions within and outside the academy.

I am attempting to highlight some important connections between past and current socioeconomic structures as they relate to the research, teaching, and function of the biological sciences by focusing on agricultural sciences. Compared to other scientific disciplines, outside of medicine, I felt there was a gap in how people were critical of racist and exploitative societal implications within the area of biology. Despite my focus on agricultural sciences and use of examples within it to ground analysis, much of this can apply broadly to scientific research and education in some form.

I interpreted this scholarship from a multitude of identities: a queer Black/New Afrikan man who grew up in the southern United States (US) countryside in a working-class family, became a first-generation undergraduate and graduate student, and is now both a scientific researcher and educator with expertise in biological sciences. My doctoral studies also occurred over a period of time that saw multiple notable events which influenced and exposed aspects of society at all levels: increased clarity of the effects of the progressing climate crisis, the global COVID-19 pandemic, uprisings against ongoing racial violence and genocide within the United States and across the globe, landmark developments in labor relations within United States higher education, and sharp increases in polarization within United States politics.

The combination of my identities, these events, and those I am blessed to be in community with were essential in uncovering the contradictions and guiding the thought process which underlies my analysis. I hope this context aids you as a reader to situate the material that is discussed here. Perhaps the most important lesson I've learned as a scientist and educator is that context is key and we're often missing pieces we aren't even aware of.

1.1 Abstract

Analyzing the socioeconomic institution of modern Science, through a focus on agricultural and plant sciences, from the conditions leading to the Scientific Revolution in the 16th century to modern day reveals the conserved influence of western European colonial-imperialism on global agricultural production. From its inception, the institution of Science is shown to be integral to the expansion and maintenance of Western imperial powers materially and socially, by driving critical revenue generation via the enabling and adoption of cash crop agriculture and through controlling the value and direction of intellectual pursuits. From the 19th century we see Science increasingly

entangled with emerging capitalist corporate and state interests, which further entrench practices that began with cash crop agriculture and prove to be detrimental to environmental health while distancing crop production from nutritional needs. I describe how these relationships ultimately limit the resources we have available for research in the modern day, hindering our abilities to address the myriad socio-scientific challenges of this century. This clarity is important when making the critical and strategic assessments necessary to direct on-going and future research and teaching efforts.

1.2 Introduction: Science is entangled with oppressive and extractive Western ideals

As scientists, we are taught and trained to observe the subject of our research with a critical eye. Operating through interpretations of the scientific method, generally the idea is that we move through the following steps in some kind of iterative fashion: pull apart all the interacting pieces, draft a theory for how they may influence each other, move to provide evidence relating to this through rigorous experimentation and data analysis, re-evaluate our theory and observations, and share findings to the wider scientific community for review. The findings may be mundane, shift our perspective of reality in drastic ways, or lie somewhere in the middle.

Much of the modern world exists as such due in large part to a rich history of scientific contributions across a wide range of disciplines and cultures as well as the teaching of that knowledge for new generations to build upon. Looking around us, the effectiveness of this critical thinking practice is clear, resulting in changes to our mental and material reality that may have sounded fictional a few generations prior. The practice of scientific research, or rather forms of knowledge-making based in repeated observation and analysis which resemble the scientific method as it is known today, has been refined and subjected to new resources over the course of human history; from its earliest inception across the cradles of civilization to the modern, globalized institution we know it as today.(1–4)

This is what I reference when discussing Science[1], which began in Europe around the 16th century during a period known as the Scientific Revolution**.** This era saw the revival of scholarly thought, an increased focus on empirical analysis formalized through the development of the Baconian scientific method, and the beginning of formal scientific institutions such as the Royal Society in the Western[2] world.(5,6) From here onward we see many rapid and revolutionary advancements that are considered essential to scientific thought — the advent of Newtonian physics, chemistry, and the Linnaeus

[1] Borrowing from Max Liborion's method of referencing the social institutions and ideals of certain words as a proper noun, capital "Science" refers to the accumulative social and economic elements of the past five centuries combining to form how we engage with and are influenced by science broadly. The practice of scientific study itself is referred to as lowercase "science".(10)

[2] “Western” primarily refers to various nations and states in the regions of Australasia, Western Europe, and Northern America, and to a lesser extent Eastern Europe and some parts of Latin America. Encompasses those where civilization or culture is predominantly Greco-Roman and Christian in origin, though it is an evolving concept with many facets and not a rigid region with fixed people or borders.

system of classifying organisms are just a few examples— and the solidification of Science as the primary avenue of knowledge and knowledge making globally.(7–9)

It is important that we view Science with the same critical thinking that is applied to research topics within any discipline, pulling apart the interacting pieces to understand how they influence each other and the implications of those revelations. When we begin to better understand how Science functions in our current society we can be honest about how it guides our practice of science, from bleeding edge research outcomes to how we teach, and the impacts of these influences.

In this text I seek to detail how Science has entangled with socioeconomic forces over the last five centuries in ways that reinforce modern inequities and crises regardless of initial intention. This occurs at all levels of the scientific process, through censorship of both the material and intellectual resources which are available for research and education. In other words, I am trying to provide a peek at the box — perhaps better referred to as a cage — which we find ourselves operating within when doing science.

1.3 Inception of Modern Science

Science is a socioeconomic institution, simultaneously built by and informing the thoughts, practices, and theory of its participants and cultural context. Each layer of its foundation affects future ones to varying magnitudes, so seeking to understand the historical conditions of its origin and establishment can provide clarity to the involved current dynamics and future predictions. Many world-defining events happened just before, alongside, and after the Scientific Revolution which entangled with the development of Science in significant ways. The earliest stages of this era were brought about by the Renaissance, a period of material and intellectual growth in Europe facilitated by a shift to the economic system of mercantilism — which prioritizes maximizing exports and minimizing imports in order to accumulate wealth — and expansion of European trade networks through imperial colonialism.[3](11,12) This expansion, later termed the Age of Discovery, began before the formally accepted beginning of the Scientific Revolution[4]. Notable events, such as the Portuguese charting naval routes to Western Africa and the Indian Ocean as well as Christopher Columbus bringing word of the ‘New World’ back to Europe, opened the door for the massive expansion of Western European kingdoms into global empires. (9)

Initial motivations behind the Age of Discovery were a mix of economic, religious, and social elements heavily influenced by the ideal of what would become known as white supremacy.[5](13) We see the legal and religious sanctioning of violent Christian European imperial expansion through the *Inter caetera* issued by Pope Alexander the VI

[3] Definition: the socioeconomic and political control and domination of a people by a nation outside of their own. Often initiated through military action.

[4] Generally believed to begin in 1543. The books *De humani corporis fabrica* by Andreas Vesalius, and also *De Revolutionibus*, by the astronomer Nicolaus Copernicus set foundations for the departure from previous scientific philosophies and methods.(5)

[5] The belief that white people constitute a superior race and should therefore dominate society, typically to the exclusion or detriment of other racial and ethnic groups.

as well as the earliest beginnings of the Atlantic Slave Trade at the very end of the 15th century.(14,15) This normalized colonial subjugation for the sake of generating wealth and influence among European powers seeking to build into empires. Then, as the Scientific Revolution began several decades later, Science itself would become an important method through which Western empires sustained themselves and expanded power across the globe. This created a feedback loop, where the accumulation of material and intellectual resources via imperial colonialism enabled massive investments into scientific research which could support the further growth of imperial power.

For example, European empires utilized Science to facilitate the expansion of their material influence through technical innovation. Advancements in math, chemistry, and physics led to significant naval and military developments which aided the establishment of colonies, extraction of their resources, and later set the stage for the Industrial Revolution.(16,17) Biology, through medicine and agriculture especially, was also crucial to supporting profit generation by enabling colonial occupation despite disease pressures; driving demand for chattel slavery and establishing large-scale cash crop plantation agriculture as a dominant system of production.(17–19)

Science also became a site of intellectual influence by rationalizing imperial expansion, positioning itself as the “correct” way to view the world in order to devalue or erase other avenues of knowledge making and knowing, and using this position to control future development of knowledge.(7) White supremacy sought to use Science to support the idea of white people being atop a hierarchy of racial classification denoting inherent biological and intellectual traits, though this work became discredited by the late 19th century.(13) This was contributed to by revision or ignorance of the actual historical trajectory of scientific research. Despite many foundational components of modern mathematics, biology, astrology, and physics having origins across a multitude of cultures, the prevailing understanding of historical scientific contribution has been and is nearly exclusive to European cultures.(8,20) This hierarchy and revision, alongside the economic and technical developments supported by imperial colonialism, provided Science the means to become viewed as an objectively beneficial and superior practice, even while aligned to wealthy Western class interests at the expense of other demographics.(7,13)

To clarify this statement with specific examples, I will explore the historical and contemporary trajectories of two phenomena: the role of agricultural sciences in the establishment of plantation agriculture and its transformation into industrial agriculture, and the establishment of pedagogical and incentive structures by Western academic institutions which shape valued scientific thought and research according to Western ideals.

1.4 Botany and Agriculture as a Tool for Empire

1.4.1 Historical Origins: Cash Crop Plantations

Within the Scientific Revolution botany saw an increasing focus on systematically categorizing plants, the environment they inhabited, and their potential attributes

(medicinal, aesthetic, agricultural, etc.).(9) Paired with Europe's burgeoning presence and economic interests across the globe, this meant botanists were doing more than introducing new additions to royal gardens. Their efforts led to new crops — such as maize, pepper, potato, and tomato — becoming ubiquitous in the Old World within a couple of centuries after being identified, exported, and bred to grow better in Europe during the Columbian Exchange, one of the largest transfers of plant species in human history.(9,21)

The 'New World' plant tobacco joined others that already had immense economic value as cash crops — such as sugarcane, cocoa, coffee, cotton, and tea — and were thus core elements of imperial economics. These crops were grown in large scale monoculture-based plantations across the globe in European colonies for export, with the New World and South Asia as epicenters.(9) The knowledge required to effectively grow, maintain, and process these crops — none of which were native to or grown extensively in Europe — in such a system was cultivated through extensive efforts by botanists and agronomists.(9) These cash crops all demanded intensive labor, required large amounts of nutrients to grow to their full potential, and were grown to be sold in increasingly globalized markets which distanced crop production from the local nutritional demands.(9,22,23) This had a two-fold effect that defined plantation agriculture and set in motion massive social and geological changes which still reverberate today.

First, labor requirements paired with the desire to maximize potential profit became the main fuel for the enslavement and transport of millions of Africans to the Americas and into the bondage of chattel slavery.(9,24–26) The large quantities of land required to grow cash crops and the economic benefits they brought to colonies contributed to the dispossession of land from and genocide of Indigenous Americans.(9,27) During these atrocities, Science co-opted and erased the contributions of many elements of Indigenous American and African knowledge pertaining to agriculture and ecology, provided it aided Western interests.(7,28) It also utilized the infrastructure of the slave trade and plantations to provide opportunities for collecting samples and data for study.(29) Other aspects of Science — now mostly discredited in modern practice — served to rationalize this by refining the concept of race and racial hierarchy.(13) Many others have detailed the multi-layered physio-social-psychological impacts these events continue to exert on the African diaspora and Indigenous American communities today.(30–36)

Second, agricultural practices on cash crop plantations were developed to maximize the amount of product exported every growing season. This led to a general lack of restorative land management practices seen in other agricultural systems of the time, such as the Norfolk four-course rotation utilized in Europe.(37) The dense monocultures often used on plantations heightened disease incidence and soil degradation, eventually leading to adoption of significant pesticide, nitrogen, and phosphorus fertilizer usage.(9,27,38) By the end of the 19th century cotton, sugar, and coffee plantations constituted over 7 million hectares, illustrating the extent to which these management practices were implemented and foreshadowing their impact.(24,39,40) This can be

considered an inflection point in the destabilization of global nitrogen and phosphorus cycles via anthropogenic sources; both of which have progressed to pressing environmental health challenges today.(37,41)

These developments — chattel slavery, Indigenous displacement and genocide, and land management practices shifting to the use of increasing amounts of fertilizer and pesticides — became profitable industries for European empires in their own right.(37,42) However, cash crop agriculture remained essential for their economic operations and was so lucrative that single crops began to comprise significant portions of the global market. Sugar alone constituted a third of the European economy by the 1700s and remained one of the most profitable industries in the world into the 1800s; by the mid-1800s cotton represented approximately half the value of U.S. exports.(9,24,25) This combination of cash crops and enslaved labor remained an incredibly important part of global economics well past the abolition of slavery across Western society. After which various forms of wage slavery that still targeted African and Asian populations came into play, enabling this system of production well into modern times.[6](43)

These plantations were the site of numerous scientific innovations across botany, chemistry, engineering, and more.(22,23,44) It is crucial to note these innovations were created or implemented in the name of increasing productivity and profit margins with little to no care for the well-being of the workers or surrounding environment. The echoes of this continue into our modern era with the emergence of industrial agriculture as the predominant method of food production globally; where it utilizes intensive monocultures, large inputs of fertilizers and pesticides, and treats crops as commodities in a global market.

1.4.2 Green and Gene Revolutions: Industrialized Agriculture

The 20th century saw incredible scientific advancements thanks to increasing industrialization and a boom in government funding for research spurred by World War II.(45,46) Plant sciences at this time were entering an unprecedented era, statistical modeling as a means to improve breeding efforts developing disease-resistant and high-yielding varieties (HYV) saw major successes, such as hybrid[7] maize transforming agricultural productivity in the US by the 1940's.(5,47) Dramatic increases in funding and institutional support from the US, Mexico, United Nations, and non-governmental organizations like the Rockefeller Foundation following World War II spurred the development of semi-dwarf wheat cultivars in Mexico by the 1950's.(45) The increased availability of inorganic fertilizer due to the invention of the Haber-Bosch process and mechanized equipment — such as irrigation pumps and large tractors — were instrumental factors in this process; these HYV were suited to high density planting and bred to be responsive to high nutrient and water availability.(46,48) As adoption of these varieties and practices spread, additional HYV of wheat and rice were developed in Asia.(9,49)

---

[6] It could be argued that many agricultural worker's relationship to their labor in modern times functions similarly. Also, plantations still exist across Africa and South East Asia.(22)

[7] Crosses of two different cultivars benefit from hybrid vigor, where first generation offspring exceed the fitness of either parent. The effect diminishes as traits segregate in following generations.

The combination of these advancements, government funding, policy decisions, and the growing use of fertilizers, pesticides, and mechanized equipment on the farm created the conditions from which the Green Revolution rose and precipitated the global adoption of industrialized agriculture.(45,50) These innovations resulted in remarkable progress: the amount of agricultural land required for the same level of food production fell by estimates of 20 million hectares, the increase in yields facilitated the human population more than doubling over the next half-century, real food prices fell for most consumers, infant mortality rates dropped in developing countries, and many rural communities around the world were brought out of poverty.(46,50) While these improvements were not evenly distributed — notably African countries continue to benefit less from such technical innovations — they were welcome developments to the general state of agriculture at the time.(46) Norman Borlaug, Yuan Longping, and Mankombu Sambasivan Swaminathan are several scientific figureheads credited with saving billions of lives.

However, the Green Revolution has its share of consequences that must be considered to fully understand the current landscape of agricultural production. This is the time period which saw global adoption of practices which have demonstrated over time to be both energetically unsustainable and sources of large-scale environmental and ecological damage.(48) The purported economic benefits have also shown to be a more complicated story, with notable negative outcomes for some of the poorest farmers.(44,45)The new crop varieties required the use of inorganic fertilizers, pesticides, and mechanization which they had been developed alongside in order to maintain the expected yields.(50) All of which are closely tied to the extraction of crude oil and have additional downstream effects that contribute to greenhouse gas emissions[8]. This has lowered the ratio of crops produced to energy input, with energy requirements outpacing yield in growth.(51) Additionally, the new crop varieties grew in intense monocultures and selection for photoperiod insensitivity meant that multiple rounds of a crop could be grown during a given year, further accelerating negative environmental effects associated with adoption of industrial agricultural practices: soil quality degradation, water system pollution, and decreased biodiversity. (45,48,50,52) The global adoption of industrial agriculture and HYVs, notably exacerbated by policy decisions, significantly diminished the use of regenerative agricultural practices and resulted in the loss of germplasm diversity found in abandoned traditional landraces. (48,49) Many farmers now found themselves reliant on private companies based out of the West to maintain their productivity, ultimately disadvantageous to smallholder farmers.(44,48,53)

[8] While there are conflicting studies about the degree and direction of influence the Green Revolution has had on greenhouse gas emissions, it is clear the intimate relationship between fossil fuel use and agriculture formed in this time period is a critical issue in efforts to address the climate crisis. The excessive use of inorganic nitrogen fertilizers especially have played a large role in reducing soil health and increasing nitrous oxide emissions.(54,55)

The Green Revolution also marked a significant change in the typical nutritional quality of a person's diet. The technical innovations were primarily implemented in cereal crops, leading to a global increase in their production and consumption while reducing the overall diversity of crops produced and consumed.(48,49) This decrease in diversity, combined with soil degradation and focus on breeding crops for high yield per unit area in monocultures, contributed to lower micronutrient content in the grains that now made up greater portions of the average person's diet.(45,48) In developed countries, there was also a rise in the consumption of animal products — such as meat, diary, and eggs — which led to a significant portion of grain production being devoted to animal feed.(56) Combined with market dynamics and technical innovations in cereal product development, much of agricultural crop land was no longer directly tied to human nutritional or textile requirements. This has led to the current situation where 80% of agricultural land on Earth is used for livestock production (grazing and feed production), despite meat and dairy products constituting only 17% of the global calorie supply. The half of global cropland used to grow food human consumption, however, translates into 83% of the global calorie supply and 62% of the global protein supply.(56) This illustrates a reduction in land use efficiency as it relates to human nutritional requirements, driven in part by the resources devoted to improving monoculture cereal crop production and processing through policy and research over other agricultural products.

However, the dawn of genetic engineering introduced new strategies seeking to improve yield and reduce loss. Large seed and chemical companies, with significant financial and technical resources developed for cereal crop and cotton research, were in a prime position to attempt to implement these new expensive technologies at scale. While not the first genetically engineered[9] (GE) product to arrive on the market, RoundUp Ready crops were the first group of them to become commercially successful and widely adopted. Today at least 80% of maize, cotton, and soybean grown in the US is some kind of herbicide resistant GE variety; a trend shared in many countries with industrialized agriculture. Initially, this innovation did lower labor requirements for farmers and greatly reduced the need for tilling seen in the common management practice at the time.(57) However, within two decades it became clear that RoundUp Ready varieties did not result in notable yield increase, led to dramatic increases in glyphosate-based pesticide usage, cultivated glyphosate-resistant weeds, and contributed to negative ecological and human health outcomes.(45,58–60)

The introduction of these technologies led to management practice shifts that increased planting intensity while diminishing rotation and intercropping usage.(61) Additionally, these seeds came with IP-related protections[10] that meant farmers were under legal

[9] Per USDA definition: Manipulation of an organism's genes by introducing, eliminating or rearranging specific genes using the methods of modern molecular biology, particularly those techniques referred to as recombinant DNA techniques

[10] It should be noted that many patents for RoundUp Ready varieties are close to being or have expired, the ability to enforce the legal protections is logistically challenging and varies across countries, and there are licenses some can obtain to allow for the saving and resale of the seeds. However, this has still considerably limited the practice of seed saving for these varieties.(53)

obligation to buy new seed every season if they were growing GE crops. (53) Along with the widespread adoption of hybrid grain varieties[11], this led to seed and chemical companies — many of which conglomerated or would soon conglomerate by the early 2000s — being responsible for significant yearly costs to farmers through chemical inputs, seed, and equipment.(53,62)

The combination of these rising yearly expenditures, market policies favoring industrialized production systems, and the increased treatment of crops as commodity products on a global market have led to greater portions of revenue generated from farming to consolidate with agricultural technology, food retail, and service industries. (53) This concentration of capital has placed these companies in position to exert market and political power to further their business interests, which are often disparate from those of the farmers and the actual nutritional requirements of a region. (53)

1.5 Censorship of Material Resources

In order to build economic power, Western empires set the stage for plantation agriculture to become the predominant mode of production globally. As the formal imperial age waned into the modern mix of corporate and neo-colonial entities wielding power, plantation agriculture grew alongside a suite of technical innovations birthing industrial agriculture as it is known today — a nearly ubiquitous production system connected to intricate global markets.

These entities have greatly influenced the global landscape of agricultural practice; representing a thread of immense socioeconomic influence running through centuries of human history, ultimately directing scientific resources for innovations that serve to reinforce practices supporting imperialist, colonial, and capitalist interests — creating a bias in the material resources available for research. Agricultural production systems became increasingly beholden to these interests, as opposed to being primarily concerned with meeting the nutritional requirements of nearby populations. Whether implicit or explicit, those in power have wielded Science — its reputation, resources, and personnel — to further their bottom line and political goals, bolstering plantation and industrial agriculture as systems which research efforts then accommodate to.

This influence has ultimately limited scientific research through censoring the material resources available. Centuries of established work and funding support research which conforms or contributes to the status quo of plantation-industrial agriculture. For example, from 2010-2019 US federal subsidies for corn, soybean, wheat, rice, and sugar were more than double those for all other fruits and vegetables.(63) The majority of agricultural research receives funding from and is conducted in or by Western entities, often relegating stakeholders in the Global South and other Third World countries as secondary contributors and beneficiaries.(64–67)

[11] While maize hybrids had been in adoption since the 1930s, wheat & rice hybrid varieties had reached the market by the early 2000s.(46)

The adoption of Green Revolution varieties and practices lead to the extinction of many indigenous crop varieties and production methods, which limits the germplasm and knowledge base from which researchers can draw to develop more climate-smart and regenerative agricultural systems. Additionally, their being the primary focus of research and development has caused traditional varieties of crops and plants with ecological significance or medicinal properties to lag behind in terms of quality and quantity of research.(50,68) These neglected plant varieties may hold essential pieces to many of the pressing concerns Science is hoping to address today, considering their wide range of traits desirable in regenerative agricultural practices and the nutritional deficit many people are facing due to dietary shifts caused by the impact of industrial agriculture on soil health and food diversity.(45,50) There have been recent studies demonstrating the advantages of utilizing “ancient” and “orphan” crop varieties to discover breeding targets of particular interest in this context.(69–72)

This limited material diversity and lack of support for research involving alternative production systems has constrained options to formulate and implement disruptive agricultural innovations These limitations are imposed differently depending on the setting where research is taking place, but remain present nonetheless. It is most prevalent in private sector research, where interests are directly beholden to the ability to generate profit. Recently, over half of global seed sales are held by 10 companies, with the top four accounting for ~40%.(73,74) These four (Bayer, Corteva, Syngenta, and BASF) are also significantly involved in the production and sale of pesticides and herbicides.(74) All 10 companies have established research and development divisions responsible for the creation of many currently utilized varieties and technologies on the farm, including using their economic resources to refine and implement discoveries made from academic settings and smaller companies. It is important to note how much material ultimately passes through corporations before reaching the field and that, for business reasons, it often must fit within the established production systems or otherwise support prior investments that have been made. In non-profit and academic settings, these pressures are lessened but still exert a notable influence: the germplasms available are largely biased towards varieties receptive to industrial agricultural practices, funding is often allocated to what are seen as safer potential returns on investment, and the body of research available to pull from is massively in favor of the current agricultural system.(47,48,50,53) Finally, as stated above, innovations are likely to pass through or be discarded by the scientific divisions of corporations before seeing wider adoption.

Additionally, the social elements of Science, chiefly the belief that it is an inherently objective practice and thus superior to other forms of knowledge-making and their potential insights, have facilitated this censorship through limiting the adoption of indigenous and other traditional knowledge or representing non-industrial agricultural production systems as outdated or inefficient when comparing yield and economic outcomes.(68,75) Notably, these comparisons lack other relevant contexts and fail to address industrial agriculture’s lack of plasticity and financial inaccessibility for small-scale farmers.(75)

These biases are important because they effectively censor the kind of research that can be undertaken. We cannot expect to solve systemic problems with agricultural production in a timely manner if we continue to be ignorant of the constraints defined by centuries of influence.

1.6 Conclusion

A final and critical point to address is whether the influences of Science I have detailed are inherent to the institution or practice — or if this even matters. Those with the most power in a society wielding it in their favor is a common feature of human history. It could be argued that what I describe as “Science” is simply referring to the influence of hegemonic power structures over time and place and is ultimately divorced from the practice of scientific research. While I have described Science here as an entanglement of imperialist, colonial, and capitalist interests, there are examples of entities outside of these political identities utilizing scientific research to further their own interests.(8,20) The practice of scientific research, in some shape or form, has a history extending long before and likely far after the rise and fall of Western imperial powers.(1,2,8,28) However, the current global context is irrevocably colored by such powers and Science and scientific research is no exception. From vocabulary to funding to available materials, unless you were exposed to or educated in other forms of knowledge-making, Western ideals and organizations provide a considerable portion of the pieces available to solve puzzles at hand. To even get to the point of developing a scientific question and experimental process, one must pull from resources, methodology, and underlying assumptions that have been heavily influenced by centuries of Western imperialism and capitalism.

I see it as important to recognize the reciprocal nature of the relationship between scientific research and these hegemonic structures — forming what I refer to as Science — and to call out the ways it has shaped conventions within scientific practice and the materials available to do so. Ultimately, like evolution, we are constrained by the current system we operate in, even as we establish new ones beyond those boundaries. If we remain ignorant to details that inform our context, we will fail to grasp information and opportunities that may prove critical in addressing the suite of challenges poised to define the next era of humanity.

It is also important to clarify that the outcomes I have detailed occurred amidst complicated contextual backgrounds and irrespective of the intentions of many scientists, educators, and policy makers. I do not sufficiently capture the complicated origins, intentions, or every outcome of the Green Revolution. Nor do I address the complicated and tumultuous backgrounds that molded the European societies which then stretched across the world, or the seeds of plantation-style agriculture being planted well before the rise of Western empires. We are influenced by history and socioeconomic conventions in unconscious ways, limited by the context of our surroundings, and it is common to see work done in good faith be co-opted by larger forces in unsavory ways. I seek to detail the influences of Science less as an indictment and more as an assessment of our current context.

While there has been no shortage of critics to how Science has conducted and implemented research over the centuries, we stand now with the benefit of hindsight. Given the ability to form this perspective and the clarity of our current pressing challenges, I believe we as scientists and educators in this realm have a responsibility to consider the realities of how Science has influenced agricultural materials and practices and seek to act accordingly. In order to truly solve systemic issues and make informed strategic decisions we must first be aware of the systems at play and the realities of what is at our disposal.

Chapter 2. A Journey Into Root Systems Biology - Understanding limitations facing the field and applying tools and techniques for informative and reproducible experiments

Including material from published works:
2.2 York, L. M., Cumming, J. R., Trusiak, A., Bonito, G., von Haden, A. C., Kalluri, U. C., Tiemann, L. K., Andeer, P. F., Blanc-Betes, E., Diab, J. H., Favela, A., Germon, A., Gomez-Casanovas, N., Hyde, C. A., Kent, A. D., Ko, D. K., Lamb, A., Missaoui, A. M., Northen, T. R., Pu, Y., Ragauskas A., Raglin, S., Scheller, H. V., **Washington, L**., Yang, W. H. (2022). Bioenergy Underground: Challenges and opportunities for phenotyping roots and the microbiome for sustainable bioenergy crop production. Plant Phenome Journal, 5: e20028.

2.3 Deng, S., Caddell, D. F., Xu, G., Dahlen, L., **Washington, L.**, Yang, J., & Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *The ISME Journal*, *15*(11), 3181–3194.

2.4 Zhou, A., Kirkpatrick, L. D., Ornelas, I. J., **Washington, L. J.**, Hummel, N. F. C., Gee, C. W., Tang, S. N., et al. (2023). A suite of constitutive promoters for tuning gene expression in plants. *ACS synthetic biology [electronic resource]*, *12*(5), 1533–1545.

2.5 Sosa, M. B., Leeman, J. T., **Washington, L. J.**, Scheller, H. V., & Chang, M. C. Y. (2024). Biosynthesis of Strained Amino Acids by a PLP-Dependent Enzyme through Cryptic Halogenation. *Angewandte Chemie International Edition*, e202319344.

2.0 Chapter Preface

The dominant mode of agricultural production requires high input usage (fertilizers and pesticides) and irrigation to maintain expected yields (48)**.** This system is energetically costly, responsible for extensive damage to environmental and ecological health, and contributing to and facing greater risks of failure from the progression of the climate crisis (48,51)**.** Despite historically being understudied in comparison to other aspects of plant biology, recent developments have indicated that a better understanding of root systems and how they interact with the surrounding environment can reveal breeding and engineering applications towards plants more well-suited for low-input agricultural systems and the increasing frequency of (a)biotic stresses (76–78)**.** The following four articles are published works that are the result of collaboration with numerous researchers in varying capacities and seek to contribute to improving our knowledge of plant root systems. How the studies detail or use approaches addressing the associated challenges of such investigation and the specific areas of my contribution are detailed at the beginning of each publication's section.

2.1 Abstract

Plant roots are essential to plant health and adaptation and important contributors to numerous ecological and biogeochemical processes. Despite this, they have been comparatively understudied aspects of plant biology. Recent developments have indicated that a better understanding of root systems can reveal breeding and

engineering applications towards plants more well-suited for low-input agricultural systems and frequent stresses. I highlight four co-authored publications concerning the study of root systems: Detailing the limitations and considerations of root system study as well as demonstrations of progress to address these challenges, metabarcoding combined with genomics and data analytics programs to inform the basis of rhizosphere microbiome heritability in *Sorghum bicolor*, constitutive promoters to fine tune gene expression in synthetic biology with demonstrated function in root systems, and utilizing root system imaging methods and software to inform the effect of a bacterial metabolite on plant-bacterial associations in *Arabidopsis thaliana*.

2.2 Bioenergy Underground: Challenges and opportunities for phenotyping roots and the microbiome for sustainable bioenergy crop production

2.2.1 Summary and Personal Contribution

York et al. 2022 is the result of a large collaborative effort across the Bioenergy Research Centers to review the current challenges, considerations, and examples of progress associated with studying root systems in order to help accelerate the area of research in bioenergy crop production. I contributed writing, reviewing, and editing for the review.

2.2.2 Abbreviations

BRCs - Bioenergy Research Centers
CP-MAS - cross polarization magic-angle spinning
CT - computed tomography
DBTL - design-build-test-learn
ECM - ectomycorrhizal
EcoFABs - fabricated ecosystem devices
FTIR - Fourier-transform infrared
GC-MS - gas chromatography-mass spectrometry
GHG - greenhouse gas
LC-MS/MS - liquid chromatography-tandem mass spectrometry
NMR - nuclear magnetic resonance
RSA - root system architecture
SOM - soil organic matter
SRL - specific root length

2.2.3 INTRODUCTION

As the transition away from fossil fuels and toward a biomass-based fuel economy accelerates, improving the economic viability and ecological sustainability of bioenergy cropping systems is increasingly critical. Annual seed-based biofuel feedstocks such as maize (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.], which have dominated the U.S. bioenergy portfolio to date, are high-yielding and profitable but often fail to provide substantive climate benefits relative to fossil fuels (79). As climate change mitigation has become a central tenet in the bioenergy initiative, it is imperative that bioenergy cropping systems provide net greenhouse gas (GHG) reduction through avoided emissions and increased soil carbon (C) storage (80). To lessen competition with food

agriculture, bioenergy cropping systems should also be targeted toward lands that are marginally productive for traditional agriculture, including regions that are prone to drought, landscape positions that are susceptible to excessive moisture, and systems where soil amendments are uneconomical or impractical (81).

Focusing on the root system of bioenergy crops and association with the soil microbiome has potential to accelerate crop productivity and sustainability, especially on marginal lands. Roots serve as the primary interface for water, nutrient, and energy transactions among plants, microbes, and soils. Belowground traits play central roles in mediating soil health, plant nutrient acquisition, feedstock stress resistance, and subsequent aboveground plant yields (Figure 2-1) (82–86). The rhizosphere is the region of soil influenced by roots, where a holistic framework is needed to understand the integrated dynamics of roots, exudates, soil, and the microbial community (87). Feedstock trait values optimizing biomass allocation to roots, root system architecture (RSA), root functional dynamics, and symbiotic associations reduce the need for exogenous fertilizer and water inputs (88), thus fostering productivity, resilience, and negative GHG budgets central to the effective deployment of bioenergy systems. However, root and microbiome phenotyping, or measuring, remains a significant impediment to realizing these positive effects that would optimize bioenergy cropping systems for marginal lands (Figure 2-1).

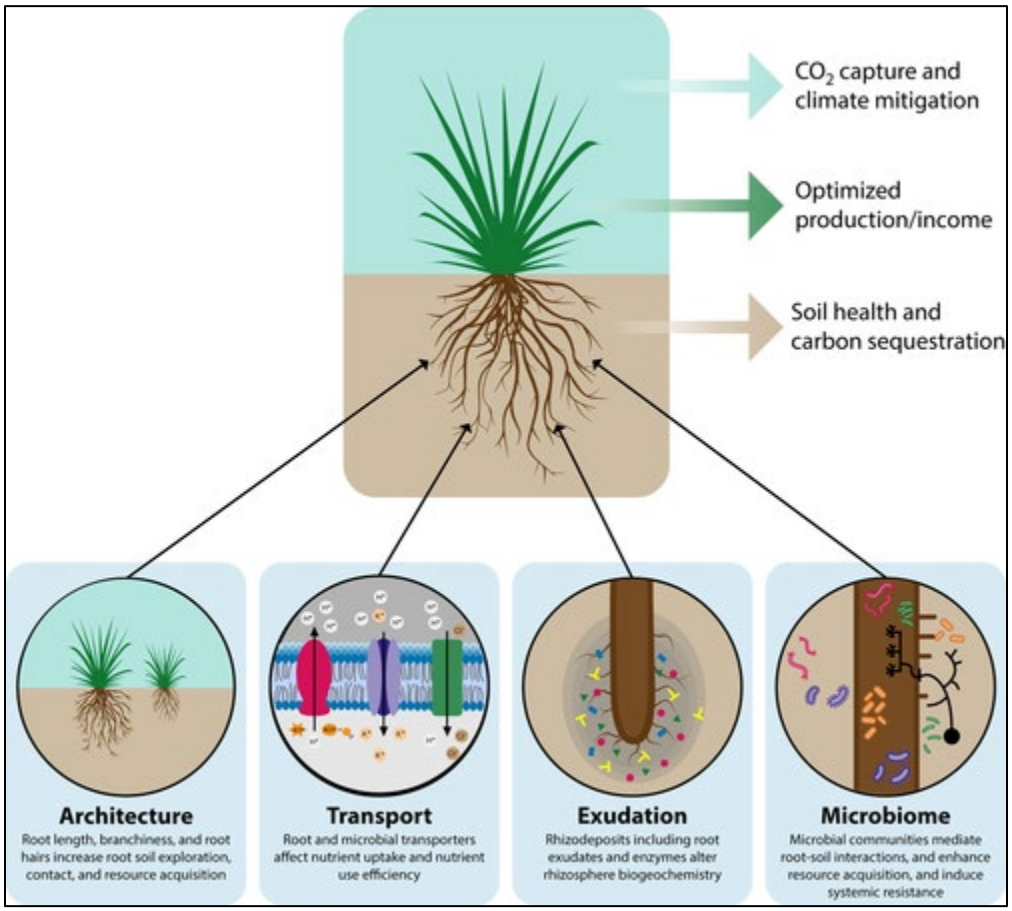


Figure 2-1. Dominant traits for optimization
Dominant traits that should be optimized belowground in bioenergy crops include those affecting root system architecture, chemistry and exudation, nutrient transport, and the microbiome. These belowground optimizations will lead to CO2 capture, increased production and income, and improved soil health and carbon sequestration

2.2.3.1 Bioenergy crops for climate mitigation

When managed properly, bioenergy cropping systems have the potential to mitigate climate change through their effect on GHG fluxes and C sequestration in soils (80). Considering that much of the aboveground plant material is harvested for feedstock production, roots are the central contributor to soil organic matter (SOM) formation and subsequent ecosystem C storage. Roots influence SOM accrual by providing fresh inputs as litter and exudates and by regulating organo-mineral complexation and soil aggregation (89,90). In fact, experimental evidence suggests that SOM accrual in perennial bioenergy systems may be more important than aboveground biomass production to offset fossil fuel use (91). Therefore, plant root trait values such as high exudation and deep rooting will increase soil C sequestration potential (92).

Roots also play a critical part in mediating nitrogen (N) transformations by rhizospheric microbes that lead to nitrous oxide (N_2O) emissions from soils. Stimulation of nitrification is primarily caused by enhanced N mineralization and availability through increased root-derived organic matter inputs, which can increase nitrate leaching and N_2O emissions (93,94). In contrast, root exudate-derived nitrification inhibitors can reduce the activity of nitrification enzymes (95), thereby reducing nitrate leaching and soil N_2O losses (96,97). In other cases, root exudates can stimulate or inhibit denitrifying bacteria, thereby either increasing or reducing N_2O emissions and nitrate leaching (98,99).

In addition to regulating N dynamics, root–soil interactions also influence the emission of methane (CH_4), particularly in waterlogged soils typical of tropical and subtropical regions where bioenergy crops, such as sugarcane (*Saccharum officinarum* L.), are grown (100). Roots can stimulate CH_4 emissions by supplying root exudates to methanogens or reduce CH_4 emissions by providing O_2 to methanotrophs in the soil surrounding the root system (101). Although there are clearly complex interactions among plant roots, microbes, and the environment, selection for plant root trait values that generally inhibit ecosystem GHG emissions and nutrient losses should be a priority for bioenergy root phenotyping efforts.

2.2.3.2 Targeting root traits for bioenergy cropping systems

Despite the obvious importance of root and microbiome traits in determining the sustainable productivity of bioenergy cropping systems, belowground traits remain poorly studied because of the spatiotemporal dynamics of soil–root–biota interactions, micro- or millimeter scale, and sampling disturbance that present methodological challenges, especially in the opaque and difficult to access soil environment. Expansion of these research efforts will require transdisciplinary collaboration to overcome current methodological hurdles. Root system plasticity, both architecturally and physiologically, is caused by soil properties, interactions with microbiota, and climatic conditions (102), which makes designing proper experimental controls challenging. Many reductive *ex*

situ approaches solve this challenge, making it possible to study root traits at very fine scales, but come up against numerous methodological inaccuracies that make the translation to field conditions uncertain. Moreover, even though the importance of deep roots has been highlighted in several reviews (103–105), the functional significance of deep roots is commonly underestimated in situ due to truncated soil sampling depths (106). Overall, the inherently destructive nature of most sampling techniques and variation in the measurement types at all scales of experimental design still represent sizable impediments to the investigation of root phenotypes (107,108).

2.2.3.3 Core Ideas

- Root and belowground traits have been overlooked in bioenergy feedstock research.
- Developing core traits and standardized protocols will facilitate bioenergy research.
- Methods exist now to facilitate belowground research for bioenergy sustainability.
- Data fusion across bio-scales needs multivariate analyses for interpretation.
- Collaboration and getting our hands dirty will be needed for the Bioenergy Underground.

Incorporating root phenotyping into bioenergy feedstock development presents an underutilized opportunity to optimize sustainable bioenergy agroecosystems. As the bioenergy movement continues forward, identifying the most economically viable and GHG-negative biofuel feedstocks will require coordinated and streamlined approaches to root phenotyping of physical, chemical, and biological traits. In this manuscript, trait will refer to a general measurable phene (e.g., root angle), while trait value will refer to a particular phene state (e.g., steep) (109). Phenotype refers to the totality of an organism's observable trait values. The process to utilize optimized belowground phenotypes will entail intensive collaboration between researchers across disciplines to overcome the methodological challenges presented by root phenotyping research. On 5 March 2021, a root phenotyping workshop was convened by the four USDOE Bioenergy Research Centers (BRCs) to initiate sharing of methodologies and discussion of future needs and collaboration regarding belowground research. To further this process, we describe pressing research questions, current challenges, and areas of opportunity for advancements in root phenotyping of bioenergy crops, applicable to other systems as well.

2.2.4 RATIONALE FOR STANDARDIZING METHODS/APPROACHES FOR BIOENERGY ROOT RESEARCH

Root traits are inherently plastic to environmental conditions and differ between annuals and perennials and between herbaceous and woody feedstocks. The lack of a standardized set of sustainability traits, as well as a lack of data integration tools to connect root traits to belowground functional processes, represent challenges to the elucidation of the role of root systems in feedstock sustainability and ecosystem function.

Obtaining a unified suite of belowground measurements for bioenergy crops would aid in addressing these challenges, further providing the foundation to develop meta-analyses of environment–species interactions and provide data streams to inform models (e.g., DayCent, Agro-IBIS, FUN-BioCROP) predicting ecosystem function. The National Science Foundation-funded NEON program (www.neonscience.org) represents an example of a data network designed to monitor the long-term health and function of ecosystems (110). Although extensive in approach, the challenges of measuring root traits have limited NEON assessment of root characters to root biomass and chemistry sampled on 5-yr interval timescales. Coordination of research activities involving belowground measures, including time- and trait-intensive measures, would provide valuable information not only on bioenergy crop influence on ecosystems, but also on the effects of longer-term environmental change and management activities on feedstock sustainability and bioenergy crop ecosystem function.

2.2.4.1 Targeting root phenotypes

Central to developing a research blueprint and set of coordinated measures is the identification of a core set of traits that reliably identify the structural and functional attributes of roots. Variation in root traits exists among genotypes of the same species and across species due to intrinsic genetic differences, and this genetic capacity can be harnessed for feedstock improvement through breeding and genetic modification. Root plasticity describes the capacity of the root system to change form and function in reaction to external conditions (102). Root plasticity itself may also be under genetic control, and although plasticity may be important for organism acclimation to the environment, researchers need to be careful assuming that any observed plasticity is necessarily beneficial (111). These core traits need to be readily measured and standardized across the research community. In addition, complementarity and integration among root traits and the microbial community are important in evaluating root response and feedstock productivity and function in varying environments (109,112).

Recent advances in imaging, image processing, and micrometeorological sensor technologies have generated an increasing number of highly multidimensional datasets alleviating previous limitations in large-scale phenotypic data acquisition that constrained breeding advances (113). Functional root traits can be targeted to reduce energy requirements (114,115), improve crop tolerance to environmental stressors (116–119), improve yields (120), and reduce environmental effects of production, all critical aspects of the sustainability of bioenergy feedstocks.

Direct measures of roots are not the only traits relevant to feedstock development, however. Indirect effects of roots on biogeochemical cycles and other aspects of the surrounding ecosystem are important and in need of standardization as well. Previously, rhizosphere has been proposed as an extended phenotype of the root (87). Even measures as relatively fundamental as respiration of roots and heterotrophs in soil are not reported in standardized formats across studies and systems. Adopting standard methods for both sampling and measurement will allow easier quantitative comparisons between field sites or entirely different biomes (121,122). The standardization of procedures embedded in the acquisition and analytical pipelines of belowground data

would facilitate cross-comparisons among studies, improve downstream analysis of data, and accelerate the development of the next generation of multiscale simulation models.

Several existing databases provide guidance for trait selection, including the TRY database (123) and Fine-Root Ecology Database (124). In TRY, aspects of root system function can be gleaned from measures of fine root branching intensity, fine root dry mass per ground area, and fine root length per soil volume (root length density). TRY also contains accessions on root chemical and physiological traits, such as fine root calcium content per fine root dry mass, fine root phenols content per fine root dry mass, and fine root phosphatase activity per fine root dry mass. These databases can serve as a starting point for trait selection and provide guidance for cataloging root traits valuable for sustainability researchers.

2.2.4.2 Root traits for selection

The adoption of core root traits and their measurement across environments and genotypes will ultimately provide targets for breeding and genetic modification. While phenotypic targets for feedstock improvement have been identified for aboveground traits such as water-use efficiency and yield (125,126), targets for important belowground traits such as rooting density, exudation, or other traits associated with resilience, productivity, and C sequestration of feedstock ecosystems have not yet been highlighted. The reliance on quantitative trait loci and genome-wide association study approaches requires an extensive evaluation of traits across broad feedstock genetic stocks. These approaches, central to the identification of loci and single-nucleotide polymorphisms conferring trait attributes, require numerous phenotypic observations across diverse genotypes and, if seeking environmental plasticity, across varying environmental conditions. This data collection intensity has been an impediment to understanding the role of root traits in feedstock performance. The development of high-throughput phenotypic platforms and advanced analytical capabilities, such as artificial intelligence, will pave the way for the next green revolution of targeted breeding (127). Functional phenomics offer a relevant research paradigm that uses phenotypic diversity and large multivariate datasets to both address genetic mapping and ecophysiological understanding using multivariate data analysis, detailed physiological studies, and simulation modeling in a pipeline that includes both hypothesis generation and testing (128).

As it stands, gaining a functional and predictive understanding of belowground root mediated traits that enable the growth of bioenergy crops in marginal lands and improve the sustainability of these crops is lacking. Leveraging collaboration across BRCs and across the entire root research community could provide potential solutions to the challenges in studying these complex belowground traits critical to ecosystem processes. Large-scale replication across a set of differing soil types can allow us to determine the stochastic and consistent root phenotypic responses. Fundamentally, BRCs share primary strategies of sustainability, feedstock development, deconstruction/separation, and conversion. These strategies could be the foundation to build upon a set of standardized experiments, results for comparison, and replication to capture the plasticity that roots display in an altered environment.

2.2.5 ROOT SYSTEM TRAITS AND FUNCTION

The efficacy of root systems in resource acquisition to support biofeedstock production and their capacity for C sequestration in soils will be reflected in a suite of traits from the whole plant, to the cellular, to the biochemical levels. These traits, root system architecture, root anatomy, physiology, and rhizosheath development, may be modified by plant response to environmental conditions and to the soil microbial community and may vary broadly by species and among genotypes within feedstock species.

2.2.5.1 RSA

Root system architecture refers to the three-dimensional spatial deployment of roots over time (129) and is influenced by plant developmental processes and genetics but also by environmental stimuli. Root plasticity is frequently observed as proliferation in localized patches of nutrients or due to soil physical constraints (111). Developmentally, RSA is largely determined by the dominant root classes, as defined by the taxonomy of the International Society for Root Research (130). All plants growing from a seed have a tap root that emerges first, and which can produce lateral roots that can further branch to have more laterals. Although this structure dominates in dicot plants, the grasses also form shoot-borne roots that can emerge from leaf nodes. Grass RSA is further complicated by tillering and rhizomes that lead to highly complicated root systems, especially in perennial grasses. Because all plants from seeds grow tap roots, and many types of roots are small diameter and fiber-like, we recommend avoiding ambiguous terms like tap-rooted or fibrous to describe root systems.

The RSAs of perennial woody feedstock species differ from those of herbaceous plants in the development of large-diameter, lignified roots formed for structural support and resource storage. These perennial structures are the site of fine root production—roots involved in resource acquisition—and the vagaries of measuring root topology and geometry have limited the understanding of the function of woody-plant root systems (131,132). These fine, first-order (using stream taxonomy) roots are generally annual in nature and vary extensively between species in structural and chemical characteristics (131,132). Effective foraging space has been defined as the soil volume in which 80% of root activity occurs and may be an aggregate measure useful for woody species (133). Since effective foraging space is functional in nature, it is a valuable concept when considering feedstock production on marginal soils or for soil C sequestration by feedstocks.

Regardless of species, the lateral roots are the sites of many of the main activities considered in feedstock production: metabolic activity, water and nutrient acquisition, mycorrhizal colonization and microbiome community activity, and rhizodeposition (131,134), thus making their assessment of central importance in bioenergy feedstock research.

2.2.5.2 Root morphology, anatomy, and physiology

Root anatomical traits, the cross-sectional structure of the root tissues, may play important roles in feedstock productivity, especially on marginal soils. Overall root diameter, root hair density and length, and the characteristics of the xylem influence water and nutrient uptake, the development of rhizosheaths, and connectivity between

soil and root (135–138). Root hairs, for example, increase the effective surface area of the root, and variation in root hair density and length may represent adaptive traits that will contribute to resource acquisition (137,138). Root hairs are important in soil water extraction as they increase soil contact, increase the effective root diameter, and reduce the diffusive limits to soil water flux (137), although the significance of root hairs cannot be generalized (139), and should be evaluated for feedstock species/genotypes of interest. The role of root hairs in nutrient acquisition and soil nutrient extraction is better defined. The ability of root hairs to bridge root–soil surfaces, penetrate small diameter soil pores, and increase the effective root diameter increases the uptake of nutrients, especially those of low diffusivity, such as phosphate (140–142).

Variation in xylem number, diameter, and length influence root hydraulic conductance, and extensive variation occurs between and within species as well as with a plant and, in the case of woody species, during the root's lifespan (143,144). Such differences in xylem anatomy represent functional adaptations to water availability (145) and may represent significant targets for selection for feedstock production under water limitation. Many cross-sectional anatomical traits related to the area and activity of the cortex are related to the concept of metabolic burden, by which C use of roots can be reduced (146).

Another, more direct measure of metabolic burden is root respiration as CO_2 release. Recently, a high-throughput method was developed for measuring specific root respiration and used for the first genetic analysis of this potentially valuable trait (115). Another physiological trait of direct relevance for roots and sustainability is nutrient uptake, which can be measured as the rate of depletion of a nutrient from solution (147). A new medium-throughput assay for the uptake of multiple nutrients indicated that substantial heritable variation existed within maize that can be harnessed for direct breeding or genetic modification (148). Physiological traits are a frontier of untapped potential for use of root traits for bioenergy sustainability.

2.2.5.3 Rhizosheaths

Rhizosheaths consist of soil particles held together by root hairs, mucilage, and other exudates produced by the root and associated rhizosphere microbial community (149). Rhizosheaths represent a bridge between the root surface and surrounding soil and play roles in water and nutrient acquisition by providing connectivity of flow paths from the soil. Often produced by cereals and grasses in response to soil stress, they are also a component of other species' root systems as well and vary among genotypes.

There are many factors involved in rhizosheath formation, including root hair length and density (150), root and microbial activity (151), and soil moisture and texture (152,153). Comparison of five switchgrass ecotypes exposed to drought stress showed a greater rhizosheath weight of ‘Alamo’ compared with ‘Kanlow’ and other ecotypes, and this was correlated with the presence of denser root and longer root hairs in Alamo (153).

2.2.5.4 Root systems of bioenergy crops

As root characteristics are central traits that support resource acquisition, vary extensively within and between species, and are influenced by the environment and

microbial symbionts, here we highlight some of the important root system features of several bioenergy crops being researched by BRCs.

2.2.5.5 Sorghum root system

Sorghum [*Sorghum bicolor* (L.) Moench] is a perennial C4 grass that is grown annually for grain production, animal forage, and bioenergy feedstock production. Energy sorghum, a photoperiod sensitive cultivar, grows more than 3 m tall with a maximal rooting depth surpassing 3 m, while producing 10–20 Mg ha_{-1} of aboveground biomass on marginal soils, and 7 Mg ha_{-1} of belowground biomass (154–156).

The tap root emerges from the seed to form the primary root system, which remains dominant for the first few weeks postemergence. Sorghum does not produce secondary seminal roots from the embryo, as closely related relatives do (such as maize). The secondary root system consists of shoot-borne roots that emerge from leaf nodes over time and that can proliferate large numbers of lateral roots. Nodal roots from the first five to seven nonelongating internodes are described as crown roots. Nodal roots from higher nodes that reach the ground are termed brace roots, and roots from even higher nodes that do not reach the ground are termed aerial roots. All nodal roots produce a heterogenous viscoelastic polysaccharide mucilage that can be colonized by N-fixing bacteria.

2.2.5.6 Switchgrass root system

In contrast to annual energy sorghum, switchgrass (*Panicum virgatum* L.) is a perennial bioenergy feedstock species. Switchgrass grown from seed starts with a single shoot and tap root system, but that forms shoot-borne roots, rhizomes, and rhizome-borne shoots and roots over time, leading to a dense root system of fine roots. However, the shoot-borne roots of switchgrass are noticeably thicker compared with other grasses. In the upland cultivars ‘Sunburst’ and ‘Dacotah’ grown over 3 yr in North Dakota, the root system to a soil depth of 110 cm accounted for nearly 27% of the total plant biomass and up to 84% with the crown (157). In perennial grasses, the crown represents the basal area of the plant where the stems and roots are connected. New tillers and roots start from the crown.

Differences in root characteristics were reported among switchgrass types with the upland types having a higher root density than the lowland types (158). Comparison of root architecture among different switchgrass cultivars showed a variation in the range of 150% in the specific root length (SRL; length per gram dry weight), with the lowland types Alamo and Kanlow having coarser roots compared with the upland cultivars ‘Forestburg’, ‘Carthage’, and Sunburst (158,159). The root hair length and density of Alamo were found to be greater than those of Kanlow (153). Profiles of root distribution in plot studies across eight locations showed that root mass averaged approximately two-thirds of the annual harvest of aboveground biomass and that approximately 50 and 75% of switchgrass roots in the top 90 cm of soil can be found in the top 15 and 30 cm of the soil profile, respectively (160). Deeper rooting in switchgrass was found to be associated with greater N uptake using stable isotopes across two ecotypes grown in both low N and water tall mesocosms (148).

2.2.5.7 Poplar root system

Considering its central role as a woody biofuels feedstock, it is surprising that so little work has focused on root system architecture in the genus *Populus*. One challenge, of course, is researching such a large and woody root system. Another issue stems from the two propagation approaches that may influence studies of root architecture: seedling vs. cutting. While evaluation of seedling root system architecture and root response to environmental factors is similar to approaches taken with *Arabidopsis* (161,162), feedstock propagation is typically via cuttings, which will be the focus here.

Root mass and extent are all strongly correlated with the size of the cutting used to establish the tree (163). In addition, extensive clonal variation in root number, length, and mass suggest lines could be selected for feedstock production on marginal soils (163–165). This variation has additionally been used to help identify genes associated with rooting propensity (166,167), establishing pathways for selection of root traits.

2.2.5.8 The response of root system architecture to abiotic stresses

Plant roots exposed to limiting soil moisture, nutrients, and other soil factors respond by modulating root traits, such as branching and root length, to maximize the acquisition of resources (168,169). A study of maize recombinant inbred lines comparing plants with contrasting lateral root number and length grown under water stress showed that the lines with fewer but longer roots had 51–67% greater shoot biomass at flowering and 144% greater yield than lines with many short roots (170). Although these results reflect intrinsic genotypic differences, they suggest plastic responses that increase root branching density may improve drought tolerance. Plasticity in the branching ability and depth of the root system has also been described in rice (*Oryza sativa* L.) in response to drought. As with maize, rice genotypes having the ability to develop deep roots were more drought tolerant than those lacking deep roots (171,172).

Root architecture in *Populus* is also environment dependent. In drought studies, root development was dependent on soil saturation (165), and lateral root growth was inhibited by PEG exposure, which may be controlled by the *PtabZIP1-like* (166) and *PtaJAZ3* and *PtaRAP2.6* (166,167) genes. Root biomass density, length, surface area, and SRL were also stimulated by N fertilization (173), and SRL and fine:coarse root ratio were strongly influenced by the form of N used (173,174), reflecting the plasticity of root architectural traits in *Populus*. Understanding the genes and networks underlying root architectural change to the environment will allow for future selection for stress resistance.

Rhizosheaths, often correlated with root hair length, increase soil-root contact and increase effective root diameter, and variation is rhizosheath development may be utilized to identify the genetic underpinnings of root characteristics that could be used to increase feedstock water- and nutrient-acquisition efficiency (175–177). Rhizosheath formation in switchgrass is stimulated by water limitation, with the Alamo ecotype exhibiting the greatest production, and this increase in rhizosheath formation was accompanied by significant increases in root hair length and density and root sugar and

amino acid concentrations (178). Clearly, root rhizosheath formation represents a potentially fruitful target for feedstock development.

2.2.5.9 Phenotyping root system architecture

Phenotyping methods for RSA can be divided into destructive and nondestructive methods and have been exhaustively reviewed (179,180). Destructive methods are the standard for root measurements to allow direct access to entire roots. These methods usually require at least destroying a portion of the root system, such as with root ingrowth cores (181) and soil coring (182) but can also involve the destructive harvest of the entire root system, such as from pot studies (148). In all cases, roots are washed free from debris and scanned on flatbed scanners that produce high contrast images for further analyses described below. Root crown phenotyping (or “shovelomics,”; (183) is another destructive method during which the top portion of root systems (root crowns) are extracted using a shovel, cleaned of debris, and the root crown is imaged for further analysis (184). These types of destructive methods are by far the most prevalent in the literature, but the nondestructive methods remain a critical need for time-series studies of growth.

Minirhizotrons and rhizoboxes are used to observe and image roots. They both rely on roots that are embedded in soil and grow along transparent surfaces. Minirhizotrons are transparent acrylic tubes that are installed at an angle into soil in the field such that specialized cameras may be inserted to acquire images along the face of the tube. In this way, root dynamics can be observed in the field, including root birth and death. For greenhouse studies, rhizoboxes are rectangular structures with usually a thin layer of soil with a clear acrylic window and are placed at an angle such that roots may grow along the surface of the window for observation. In the past, manual tracing and tracking was used (185), but recent advances in image analysis may allow striking progress in their widespread adoption (186,187).

Rhizotron and rhizobox methods rely on roots growing along artificial surfaces for observation, which are known to have an influence on root growth. Therefore, a fascinating frontier of root research is the use of various electromagnetic signals that can penetrate soil and nondestructively reveal roots, such as nuclear magnetic resonance (NMR), X-ray computed tomography (CT), and radar. Using NMR, the water in and around plant roots is mapped to create two- or three-dimensional images of the root systems. High-field NMR can image root systems in high detail, but structured soil provides background interference (188). Low-field NMR can image root systems in similar detail, but it is less affected by paramagnetic elements in the soil, thus allowing for imaging of root systems in a wide range of natural soil types (189). This technique provides a nondestructive way to observe the root system of plants, however it is currently time-intensive, requires the use of soil cores or a rhizotron system, and has a minimal resolution of ~1 mm, meaning fine roots may be difficult to resolve. X-ray CT uses X-rays from multiple angles to compile 3D volumes of root systems embedded in soil (190). Currently, microCT systems can generate images with pixels as small as 100 nm, relevant for fine roots; however, the nature of X-ray attenuation means that roots may be difficult to segment from the soil because of similar intensity. Nuclear magnetic

resonance and X-ray represent the frontier for detailed reconstructions of 3D root systems over time, but these approaches are presently constrained by being mostly limited to laboratory use and imaging columns or pots.

For field use, ground penetrating radar is the most studied nondestructive method for roots, during which electromagnetic pulses are broadcast into the ground (191). The pulses scatter back to the instrument receiver when they cross boundaries of contrasting materials, where the received information can be used for 3D reconstruction. This nondestructive technique can be applied to large areas of a field for high-throughput phenotyping. However, this technique is unable to reconstruct finer root details currently, so it is best used for larger belowground structures such as in tuber crops, trees, and nodal roots. Another promising approach for radar may be not attempting 3D reconstruction, but rather using the signals directly for estimating root length density, which may also predict length of finer roots in aggregate (192). Various electrical methods have also been proposed but generally require sensors to be placed in the ground or on the plant (193).

2.2.5.10 Root system architecture image analysis

Roots are imaged across a wide degree of modalities, such as in situ with minirhizotrons or soil pits, in rhizoboxes, on colored backgrounds, or with flatbed scanners (180). Therefore, image analysis approaches are often specific to particular types of collected images. Software that works with 2D branched and connected root system images such as root crowns or seedling root systems grown on blue paper included EZ-Rhizo (194), SmartRoot (195), RootNav (196), and ARIA (197). In many cases, roots are not connected as they have been excavated from the field or pots, and these roots are typically imaged on a flatbed scanner then analyzed using the commercial WinRhizo software (Regent Instrument Inc.). However, the previous *de facto* standard for root measurements, WinRhizo, was shown to drastically underestimate root volume when root diameters within an image vary because the software uses a volume formula using total root length and the average root diameter (198). Most recently, RhizoVision Explorer was released as a ready-to-run executable for Windows that facilitates interactive root image analysis as well as batch analysis (199). RhizoVision Explorer correctly calculates root volume, and analyzes both connected root systems, such as seedlings, as well as disconnected roots washed from soil cores or pots. However, RhizoVision Explorer and WinRhizo require high contrast root images for successful root identification. RootPainter (186) uses machine learning after initial training to segment roots from more complex backgrounds, such as in minirhizotrons or rhizoboxes, and these segmented root images can then be used with RhizoVision Explorer. The combination of RootPainter, or similar tools, for segmenting complex imagery with RhizoVision Explorer for feature extraction is a promising pipeline for future root research.

The frontiers of RSA phenotyping include broadening the base of scientists who are root-enabled by releasing open-source software tools, sharing standardized protocols, and pushing the frontiers of nondestructive methodologies. Knowledge gaps include how phenotyping roots among plant ages or environments affects results, or whether

phenotyping vegetatively propagated plants is indicative of seed-grown plants. Because root systems represent a substantial carbon expenditure for the plant, carbon is potentially diverted from the root system to achieve higher shoot biomass yields during selection. However, it is unknown if selection on aboveground biomass in bioenergy crops tends to decrease the relative allocation to the root system.

2.2.6 ROOT STRUCTURAL CHEMISTRY, METABOLITES, AND EXUDATES

In addition to root traits, root biochemical characteristics and rhizosphere modifications represent critical traits that affect both biofeedstock productivity and ecosystem function (87). These root traits vary greatly across and within species and life forms and are modified by environmental conditions (200).

2.2.6.1 Root chemistry influences soil C dynamics

Root chemistry, including broad traits such as C:N ratios and specific traits such as carbohydrates (i.e., cellulose and hemicellulose) profile and lignin structures and content, affect decomposition rates and therefore, potentially contribute to SOM stocks. For example, C:N and lignin are negatively correlated with root decomposition rates, whereas holocellulose content is positively correlated with root decomposition for perennial bioenergy grasses (201–203). The compositions of roots and its chemical features also vary depending on the root resources. Although these correlations are not universal (201), the potential of root chemistry to influence SOM is important because total root biomass can exceed shoot biomass in some species (204). Indeed, root-derived SOM may exceed that derived from shoots by 2.4 times (205,206). Further, microorganisms can process and incorporate root C into biomass four times faster than aboveground plant residues (205). Given current thinking that microbially derived necromass, metabolites, and decomposition products, rather than recalcitrant plant material, account for the majority of stabilized SOM (207–210), the faster cycling of root material through microbial biomass can lead to greater SOM accrual. Thus, root chemistry plays a significant role in influencing the process rates and decomposer community structure and, ultimately, the accrual of C by biofeedstock systems.

2.2.6.2 Root metabolites

Fundamental differences in the metabolism of bioenergy feedstock species and genotypes may reflect fundamentally different ecological capacities and/or responses to environmental factors. Plant acclimation to stress often includes specific sets of metabolic adjustments, including changes in primary and secondary metabolites as well as the establishment of reactive oxygen species-scavenging responses (211,212). While such metabolic fingerprints may be used as a trait for bioenergy feedstock selection, especially for use on marginal soils, data on plant metabolic profiles is largely limited to leaves (213–216).

Primary and secondary metabolites and levels of antioxidant compounds are often reflective of metabolic activity and stress response. Switchgrass lines differing in rhizosheath production and drought response exhibited root amino acid, sugar, and organic acid profiles that significantly changed in response to water limitation, with arginine, isoleucine, methionine, and cysteine and kestose, raffinose, fructose, fucose,

sorbose, and xylose increasing as rhizosheath size increased (153). Metabolic profiles in roots of switchgrass also change in response to nutrient stress (217). In *Populus nigra* L., water limitation altered C allocation to the roots, with concomitant changes in C partitioning between nonstructural carbohydrate pools and induction of antioxidant enzymes (218). In *Populus deltoides*, water stress differentially affected the flavonoids chrysin, myricetine, kaempferol, and isoferulic acid in roots of different genotypes and may reflect the adaptability of poplar plants to water stress (219).

In addition to influencing feedstock environmental stress resistance, such broad differences in root metabolic profiles may influence rhizodeposition (below) as well as the functional ecology of the rhizosphere (220). Thus, metabolic assessment and engineering may represent avenues to enhance feedstock sustainability and control agroecosystem GHG fluxes.

2.2.6.3 Root exudates

As another important root trait, rhizodeposits include passively released low molecular weight compounds, active exudation of secondary metabolites, proteins, and mucilage, and sloughed root cap and border cells (221). Rhizodeposition accounts for 10–40% of primary productivity (222), but of the components comprising rhizodeposits, root exudates have received the most attention. Root exudates are organic compounds including both primary (e.g., sugars, amino acids, organic acids, fatty acids) and secondary metabolites (including volatile organic compounds and plant hormones), that shape a variety of plant and soil traits ranging from root growth and architecture to soil physical and chemical properties (e.g., soil pH, nutrient content, water holding capacity) (223–227). Dynamic interactions between roots and soil microorganisms mediated by root exudates can benefit plant growth and stimulate plant water and nutrient acquisition while enhancing the resistance of plants to biotic and abiotic stressors, creating emergent properties and traits (228). Therefore, root exudates and the rhizosphere should be considered part of the root extended phenotype (87,225).

2.2.6.4 Root exudates respond to the environment

Rhizodeposition is both responsive to and alters the environment (229–231). Exudation is altered by drought, with broad changes in profiles depending on plant species, where increases may reflect active responses of plants to modify the rhizosphere or increased deposition as roots lose cell integrity and die (232). Other studies have documented exudation shifts to sugars, antioxidants, and secondary compounds under drought (233). Changes in plant nutrient status can also lead to changes in the exudation rates of various carbohydrates, organic acids, and amino acids relative to each other and in total (234–239). These changes in rhizodeposition may allow bioenergy feedstock species to engineer their soil environments, build rhizosheaths, or otherwise modify the rhizosphere to aid in the acquisition of water and limiting nutrients, with concomitant effects on ecosystem processes including C storage and nutrient cycling.

2.2.6.5 Root exudates affect the soil microbiome and vice versa

In addition to altering rhizosphere water and nutrient relations, chemical composition of root exudates can act as a selective agent on the rhizosphere microbiome, altering

microbial community structure and soil ecosystem function(227,240). Through exudates, plants can communicate with and regulate microbial communities, stimulating beneficial microbes and inhibiting the growth of pathogenic microbes (161,223,226,239). At the same time, the composition of the soil microbial community also influences rhizodeposition (241) through changes in gene expression induced in the root by symbionts/associations (162,195). Indeed, different root microbiomes often exhibit fundamentally different rhizodeposition profiles (235,242,243), which may influence feedstock productivity and sustainability on marginal soils.

In the rhizosphere, root exudates represent an easily accessible source of labile organic C. This readily available C influences microbial population size, activity, and community structure (227,240,244,245) and may also accelerate the mineralization of SOM in an effect referred to as priming (246). Much of the priming effect observed in the rhizosphere is likely the result of microbes mining SOM for N while using plant C for metabolism and biomass construction (247–250). On the face of it, this priming effect would seem to have a detrimental effect on SOM stocks, but this is usually not the case. Large quantities of root exudates may select for microbial communities dominated by fast growing, copiotrophic species (251,252). The resulting faster microbial biomass turnover of copiotrophs would increase the quantity of microbial residues generated and thus available for the formation and accrual of SOM (89,253). In addition, root exudates themselves can be rapidly incorporated into SOM as the exuded low-molecular weight compounds easily sorb to mineral surfaces (247).

As we consider root chemistry as part of the multidimensional root phenotype, there remain significant gaps in our knowledge, leaving pressing research questions. Although there is already evidence of links between root chemistry and the rhizosphere microbiome, given the importance of this microbiome for plant health and productivity our knowledge is still in its infancy, especially with respect to bioenergy cropping systems. There remain many questions surrounding mechanisms by which root exudates shape microbial community assembly and function. For example, some biological nitrification inhibition compounds are released via secondary transport (192,254). However, it is generally not known how these transport mechanisms mediate the response of root exudation to rhizosphere soil conditions. We have also observed that low soil moisture can potentially induce changes in root exudate quantity and composition due to plant stress response and can affect the movement of hydrophilic vs. hydrophobic root-derived compounds to alter the effect of root exudates on the microbial community and soil processes. Overall, we need a better understanding of the mechanisms and characteristics of root exudates, including composition and quantification of exudates, mechanisms of transport from roots, and relationships between root architecture and developmental stage and exudate location, quantity, and composition.

2.2.6.6 Root biochemistry and exudation measurement and methods

The analysis of lignin, cellulose, and hemicellulose content is most often determined using wet-analytical methods, but several studies have highlighted improvements in experimental throughput by NIR techniques (255). Lignin structural features are

commonly reported by ^{13}C quantitative NMR or heteronuclear single quantum coherence of isolated samples, although solid-state ^{13}C cross polarization magic-angle spinning (CP/MAS) has also been frequently employed (256,257). The latter approach does frequently suffer from spectral overlap, whereas heteronuclear single quantum coherence minimizes this problematic issue. Raman imaging has been used to map-out the location of lignin in the root cell wall. Fourier-transform infrared (FTIR) spectroscopy has also been used as a companion method, but the structural information is limited in comparison to NMR methodologies. A hydrolysis-ion chromatography protocol is most often used to determine carbohydrate sugar profiles in roots. New data generated broadly by the use of these methods will provide insights into how root tissue chemistry is altered as bioenergy crops are engineered for biorefining of aboveground biomass, with implications for soil C and rhizosphere microbial communities.

Given the importance of root exudates, it is critical that we use a wide-range of analytical methods to assess variation in rhizodeposition among feedstocks and the potential roles of exudation in environmental response. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) and gas chromatography-mass spectrometry (GC-MS) are leading approaches due to their ability to characterize complex mixtures with relatively high sensitivity. However, these approaches typically require liquid or solid phase extraction to fractionate metabolites from salts and other sources of interference and metabolite identifications are typically based on comparison with authentic standards. NMR, on the other-hand, can directly determine the structures of novel compounds without this reliance on standards, however, it is less suitable for complex mixtures and low abundance compounds.

While LC-MS/MS, GC-MS, and NMR are used for analyses of ex situ samples, there are a range of spectroscopic methods, especially FTIR (258,259) that can enable direct in situ analysis of chemical groups, albeit by looking at bulk spectroscopic features vs. the structures of constituent molecules. Fourier-transform infrared can also provide information on the localization of chemical classes within the rhizosphere which can be complemented by using mass spectrometry imaging to localize metabolites (260). Stable isotopes provide a complementary method to determine the temporal rates of production of exudates, for example, by introducing $^{13}CO_2$ into a gas tight growth chamber. The resulting isotopically labeled metabolites can be resolved from the ^{12}C metabolites to determine when particular metabolites are produced and even calculate exudation rates. In addition, the ^{13}C-labelled roots will greatly enhance the NMR signals thus enabling the characterization of low abundance components or metabolites.

We see several opportunities to standardize analytical methods for the analysis of exudates across institutions, especially LC-MS/MS and GC-MS analyses. First, the use of common growth conditions including experimental systems such as fabricated ecosystem devices (EcoFABs) (261) that facilitate exudate collection to help ensure comparable samples are generated. Similarly, standardization of exudate extraction and analytical methods can allow community researchers to directly compare results and ideally leverage authentic standard libraries for compound identification. Finally, by comparing results across diverse plant species we can define components of core

exudomes for bioenergy crops to define internal standards that can be used for metabolite identification and quantification. Specifically, spiking in isotopically labeled versions of core exudome metabolites can significantly improve the ability to confidently identify and quantitatively profile these metabolites. Internal standards are especially important in complex sample matrices such as exudates and rhizosphere samples that contain interferences that may suppress the ionization of metabolites of interest. Overall, by adopting common protocols for exudate collection and analysis, including internal standards, research will be accelerated by enabling scientists to compare data and build on each other's results.

2.2.7 ROOT MICROBIOME, COLONIZATION, INTERACTION, AND SIGNALING

The rhizosphere represents the critical interface between plants, microbes, and soil (262). Plants mediate and drive belowground biological interactions and activities in the rhizosphere through root exudation (263). These exudates, along with other rhizodeposits, stimulate microbial activity and help drive synergistic activities of diverse and functional groups of soil microbes, including N-fixing bacteria, P-solubilizing siderophore-producing microbes, as well as arbuscular mycorrhizal fungi and ectomycorrhizal (ECM) fungi, which help promote plant growth and resilience through provisioning plant nutrition and by priming plant immunity (264–267).

The root microbiome includes rhizospheric bacteria and fungi that are less intimately associated with roots, rhizoplane organisms that may colonize and form biofilms along all plant root surfaces, and symbiotic partners that all affect plant health and growth (268–270). Rhizoplane microbiota may survive by exchanging metabolites with the plant; however, most are not specialized to invade plant cells or apoplastic tissues (271). However, endophytic microbiota are able to evade plant defenses and live within plant tissue as symbionts (272,273). These include the well-known nodulating N-fixing bacteria, arbuscular mycorrhizal fungi and ECM fungi, as well as a diversity of other bacteria and fungi (274–277). Rhizoplane and endophytic members of the root microbiome participate in direct signaling and molecular communication with the host plant and may aid in damping stress signals, including reactive oxygen species, provision of macro- or micronutrients, or protecting the host through chemical means (278,279). Such intimate plant-microbe signaling can involve diverse molecules, including low molecular weight compounds, small secreted peptides, and lipochitooligosaccharides, altering plant development and response to the environment (280,281). Rhizoplane and endophytic microorganisms are also active in acquiring and transporting nutrients and water out of soil for their plants’ hosts, thus, they may directly and indirectly improve stress-tolerance of plants growing in marginal environments. In this way, their activities can help to contribute to ecosystem-scale sustainability benefits of bioenergy crops such as improving soil structure and C stability and reducing soil N and C emissions (282,283). Thus, belowground root-microbe biology, communication, and dynamics are fundamental to plant health and the development of sustainable bioenergy cropping systems.

Much of our understanding about root biology comes from reductive experimental systems or greenhouse experiments, which have limited translation to field scale

ecosystems. This limitation is in part due to difficulties capturing spatial and temporal variation in root systems and complicated by limited understanding of how the wide range of traits relevant to belowground processes influence each other (112). For example, both root chemistry and architecture are known to influence microbial community composition, while common microbial mutualists, such as mycorrhizal fungi, have in turn been shown to influence root chemistry and architecture, yet there is little work correlating root traits to microbiome structure and function (223,284–286).

Design, build, test, learn (DBTL) is an ideal framework for both top-down and bottom-up microbiome design in closed systems due to the regulation of environmental conditions (287). However, the success of the engineered single-species microbes within complex ecological networks, such as the rhizosphere, is likely limited due to the extensive functional redundancy, environmental variability, and emergent properties characteristic of the rhizosphere microbiome. Moreover, many endophytic and rhizosphere microorganisms have not been cultured, leaving gaps in functional profiles of plant-associated microbiota (288).

An alternative approach is to engineer microbial recruitment and recognition mechanisms into bioenergy germplasm. For example, transgenic *Arabidopsis thaliana* expressing a *Populus trichocarpa* lectin receptor-like kinase protein (PtLecRKL1) involved in the ECM symbioses, displayed ECM fungal sheaths despite *A. thaliana* being a non-ECM species (289). Lectin receptor-like kinase proteins are pattern recognition receptors that bind to microbial-associated molecular patterns and initiate *Populus* symbiosis with ECM fungi, *Laccaria bicolor* (290). These results suggest pattern recognition receptors may be a target for engineering plant symbioses (291). Unfortunately, the molecular mechanisms governing plant perception of microbial symbionts, specifically within bioenergy species, require extensive investigations. DBTL can be used to methodically dissect the genetic and regulatory elements, as well as signaling pathways involved in the specific recruitment of plant-growth-promoting rhizobacteria and fungi. Identification of the interspecies signaling involved in the recruitment of target bacterial and fungal taxa provides an opportunity to incorporate sustainable root-microbial traits into bioenergy germplasm development.

Some of these questions are being pursued amongst the BRCs, including recruitment and turnover of rhizosphere microbiomes, how their distribution and activity change with depth, and what taxa constitute core (shared) communities. In these investigations, central themes become apparent, such as seeking strategies to visualize and localize specific microbial taxa, the importance of biofilms, and utilizing mixed fungal/bacterial communities. All of these aspects are limited by currently available methods and variance in what measurements are considered relevant across the discipline.

2.2.7.1 Engineering plants for bioenergy: effects on the root microbiome

Within bioenergy science, there is a significant focus on engineering of bioenergy crops with reduced feedstock recalcitrance, which typically involves changes in cell wall composition, to facilitate conversion to fuel. Several studies have shown that changes in cell wall composition alter plant-microbe interactions and, likewise, plant-microbe

interactions cause changes in plant cell wall composition (292–294). In addition, several groups are engineering bioenergy crops that accumulate bioproducts or bioproduct precursors (e.g., for renewable and biodegradable plastics) (295–297). Lignin is an important component in these interactions and is also the target of many bioenergy crop engineering projects. However, very little is known about how the engineering for improved downstream processing affects plant resilience to the environment and interactions with microbes.

Although many projects are aimed at reducing recalcitrance, such as by reducing lignin in above-ground biomass, there is also an increasing interest in minimizing GHG emissions and increasing soil C sequestration of bioenergy crops (e.g., by increasing compounds such as lignin or suberin in roots). However, accumulation of such plant-derived compounds in the roots or soil must be studied in the context of soil processes and the rhizosphere microbiome to understand how the changes in root properties affect microbiome composition and the production of relatively stable soil C (298,299).

Bioenergy researchers are interested in whether we can engineer plants or their microbial partners to create a more adaptive root microbiome to benefit plant performance in marginal environments and soils. However, there is currently insufficient knowledge to rationally design plants for recruitment of specific microbiomes. We do anticipate, through the DBTL framework, that it will be possible to engineer plants to produce specific exudates that stimulate certain microorganisms over others to promote plant growth and soil C sequestration, while reducing GHG emissions and increasing the value of bioenergy crops.

2.2.8 SYNTHESIS

The primary focus of bioenergy feedstock improvement efforts over the past two decades has been on aboveground chemistry and productivity. The goals of sustainable bioenergy crop production and enabling a C-neutral bioeconomy will, however, require consideration of both above- and belowground plant system performance (300). Plant root systems and the surrounding soil environments that they influence, or rhizospheres, are complex, dynamic ecosystems that affect bioenergy crop yields and soil health (87). While root traits have long been considered a phenotypic target in plant breeding (301), the complexities of belowground ecosystems and the difficulties accessing them have hindered our understanding of these key environments. With dozens of reports on new root imaging technologies and image analysis software emerging from the research community in the past decade, we are collectively better poised now than ever to represent the “hidden half” in agroecosystem studies that consider the aboveground and belowground simultaneously (127) but have to separate the hype from the useful. While researchers often report that roots are the “hidden half” and difficult to study, they should understand that roots can successfully be studied for contributions to scientific knowledge using readily available methods, even while we wait for the promises from frontier technologies.

There remains a dearth in approaches that have both streamlined higher-throughput data acquisition and analysis of RSA, especially under field conditions (183,302–305).

Root phenotyping technologies offer tradeoffs from lower throughput, high-resolution to higher throughput and lower resolution, and there is a growing demand in the community for versatile and accessible technologies that are in the sweet spot of higher throughput and resolution. The broader research community has a significant role to play in driving the expectations and standards for empirical methodology, metadata and data quality and storage such that bioenergy crop data generated from across species, field sites, growth matrices, and phenotyping methodologies are conducive for robust meta-studies and modeling efforts.

Advancements in plant and root system characterization technologies are, however, creating opportunities for integrated experimentation across the scientific community that can expedite progress in multi-scale understanding on how root structural and chemical traits are influenced by plant genetics, climatic conditions, and edaphic factors. One such advancement is in the development of fabricated ecosystems that are designed to link observations across benchtop, greenhouse, and field experiments. These systems range in size, cost and complexity, from a small hand-held size, laboratory EcoFABs to those that can encompass entire facilities (e.g., EcoTrons, EcoPODs) (306–308). These engineered ecosystems could be especially useful for studying the increased below and aboveground production found in mixed assemblages of grassland species (91). Democratized technologies and access to new integrated lab and field-scale facilities as a community resource will allow for the integration of root system architecture, microbiome composition, chemical mass balance, soil matrix, and exudate measurement or sampling technologies into existing research pipelines while increasing the throughput of testing responses to abiotic (environmental) and biotic (plant and microbe) changes, thus accelerating hypothesis generation and testing (261,309).

Recent proliferation of phenotyping and measurement technologies is resulting in a dramatic increase in the number, heterogeneity, and complexity of datasets. Understanding features, spatiotemporal patterns, and correlations from large heterogeneous datasets requires use of appropriate statistical and data analysis methods. The community needs to embrace multivariate methods (310), because too often multivariate data is treated with many univariate analyses. Machine learning methods are often advocated due to their ability to find nonlinear connections in data that can offer substantial predictive value, however these methods can often be a black box that lack interpretability so crucial for increasing scientific understanding. Therefore, more exploration of interpretable machine learning (311), or explainable artificial intelligence, will be needed for the success of these proposed research programs across bio-scales.

Concerted community efforts are needed to be able to bridge bio-scales of experimental studies and maximize utility of plant phenotypic data collected towards modeling efforts. To effectively accelerate our understanding of how root architecture, rhizosphere function, soil health, and plant yields are linked to and influence one another, standardized data collection and reporting is needed. This is particularly true when trying to integrate research conducted in laboratory ecosystems with field observations

collected from minirhizotrons or imaging of root architecture of excavated plants (312,313). Considerations should be made when collecting data on the types of data (e.g., RSA), the scale of measurements, and the methods of measurement. This is particularly important to facilitate effective simulation modeling of these systems. Selection of the core traits and creation of a protocol library will be essential to the success of this program. It is encouraging to already see an organic increase in community synergy around root imaging approaches, standardizing nomenclature, and usage of root image data sharing platforms (314). The plant science community at this juncture needs similar synergy in the context of root chemistry characterization. Adoption of standard protocols for root and exudate collection and chemical analysis as well a commitment to F.A.I.R. data principles (315) in analytical chemistry data and library sharing for root, rhizosphere, and soil analyses will catalyze meaningful meta-studies and integration of independently generated datasets into models. In a broader context of root phenotyping for fundamental plant or applied food, fuel, or fiber crops research, scientific societies and funding agencies can be powerful catalysts in promoting and incentivizing coordination of research endeavors and leveraging alignment within and across funding agencies.

2.2.9 CONCLUSION

Ultimately, we envision a belowground bioenergy research community collecting data across sites and experiments to build community knowledge even when pursuing individual aims (Figure 2-2). Many core root traits can be collected now in almost any lab using destructive but simple soil coring, washing, scanning, dry weighing, and analysis with RhizoVision Explorer. These traits include root length density, specific root length, average diameter, and root tissue density, which have recently been described in a guide on the methods and theory of root trait ecology (112). In the longer term, nondestructive root phenotyping tools should continue development until they are ready for wider community adoption. Measuring core rhizosphere traits, also referred to as the rhizosphere metaphenome (316), which includes chemical, soil, and microbiome characteristics such as metabolomes, proteomes, transcriptomes, and metagenomes, as well as microbial functions and responses to the environment, will require coordinated research hubs and partnerships among bioenergy research centers. These core trait measurements will require standardized protocols and data sharing platforms. The core message of the Bioenergy Underground is that we need to get our hands dirty and get digging to realize the unprecedented opportunities to increase feedstock production on marginal lands while improving carbon sequestration to combat the climate emergency (317).

Figure 2-2. Illustration
Understanding roots, soil, and the microbiome in the context of bioenergy cropping system sustainability requires significant data fusion and multivariate analytics combining data from technologies such as UAVs (uncrewed aerial vehicles), satellites, and flux towers aboveground, and rhizotrons, minirhizotrons, and sensing technologies such as ground penetrating radar belowground

2.2.10 ACKNOWLEDGMENTS

This work was funded by the Center for Bioenergy Innovation, a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science, Oak Ridge National Laboratory is managed by UT-Battelle, LLC under Contract DE-AC05-00OR22725 with the U.S. Department of Energy to L.M.Y., J.R.C, U.C.K., Y.P., A.J.R., A.M.; 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 to A.T., G.B., A.L., L.K.T, D.K.K.; the DOE Joint BioEnergy Institute supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy to H.V.S., L.W., J.H.D.; the DOE Center for Advanced Bioenergy and Bioproducts Innovation, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-SC0018420 to A.C.V.H., E.B.B., A.F., A.G., N.G.C., C.A.H., A.K., S.R., W.H.Y.; a Laboratory Directed Research Development (LDRD) LBNL grant to J.H.D; and the Microbial Community Analysis and Functional Evaluation in Soils Program at Lawrence Berkeley National Laboratory DE-AC02-05CH11231 to T.R.N and P.F.A. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable,

worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). We would like to thank Chelsea Mamott and the GLBRC communications team for the figures.

2.2.11 AUTHOR CONTRIBUTIONS

Larry M. York: Writing-original draft; Writing-review & editing. Jonathan R. Cumming: Writing-original draft; Writing-review & editing. Adrianna Trusiak, Writing-original draft, Writing-review & editing. Gregory Bonito: Writing-original draft; Writing-review & editing. Adam C. von Haden: Writing-original draft; Writing-review & editing. Udaya C. Kalluri: Writing-original draft; Writing-review & editing. Lisa K. Tiemann: Writing-original draft; Writing-review & editing. Peter F. Andeer: Writing-original draft; Writing-review & editing. Elena Blanc-Betes: Writing-original draft, Writing-review & editing. Jonathan H. Diab: Writing-original draft; Writing-review & editing. Alonso Favela: Writing-original draft, Writing-review & editing. Amandine Germon: Writing-original draft, Writing-review & editing. Nuria Gomez-Casanovas: Writing-original draft, Writing-review & editing. Charles A. Hyde: Writing-original draft; Writing-review & editing. Angela D. Kent: Writing-original draft, Writing-review & editing. Dae Kwan Ko: Writing-original draft; Writing-review & editing. Austin Lamb: Writing-original draft, Writing-review & editing. Ali M. Missaoui: Writing-original draft, Writing-review & editing. Trent R. Northern: Writing-original draft, Writing-review & editing. Yunqiao Pu: Writing-original draft, Writing-review & editing. Arthur J. Ragauskas: Writing-original draft, Writing-review & editing. Sierra Raglin: Writing-original draft, Writing-review & editing. Henrik V. Scheller: Writing-original draft, Writing-review & editing. Lorenzo Washington: Writing-original draft, Writing-review & editing. Wendy H. Yang: Writing-original draft, Writing-review & editing.

2.3 Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome

2.3.1 Summary and Personal Contribution

Deng et al. 2021 provides an example of utilizing 16S rRNA metabarcoding alongside genomic resources to reveal a heritable basis influencing rhizosphere microbiome assembly in *Sorghum bicolor,* demonstrated in field conditions and validated at a different developmental stage in controlled settings. This illustrates the utility of compiled databases, emerging genetic techniques, and standardized methodologies for enabling novel insights. My primary contributions were protocol development, data collection, processing, and analysis for the experiment to validate findings from the field trial in a controlled growing environment as well as co-authoring the manuscript.

2.3.2 Introduction

Recent work has shown that root-associated microbial communities are in part shaped by host genetics(318–321). A study comparing the root microbiomes of a broad range of cereal crops has demonstrated a strong correlation between host genetic differences and microbiome composition (321), suggesting that a subset of the plant microbiome may be influenced by host genotype across a range of plant hosts. In maize, these genotype-sensitive, or “heritable”, microbes are phylogenetically clustered within

specific taxonomic groups (322); however, it is unclear whether the increased genotypic sensitivity in these lineages is unique to the maize microbiome or is common to other plant hosts as well.

Despite consistent evidence of the interaction between host genetics and plant microbiome composition, identifying specific genetic elements driving host-genotype dependent microbiome acquisition and assembly in plants remains a challenge. Recent efforts guided by a priori hypotheses of gene involvement have begun to dissect the impact of individual genes on microbiome composition (323,324). However, these studies are limited to a small fraction of plant genes predicted to function in microbiome-related processes. In addition, many plant traits expected to impact microbiome composition and activity, such as root exudation (325) and root system architecture (286), are inherently complex and potentially governed by a very large number of genes. For these reasons, there is a need for alternative large-scale and unbiased methods for identifying the genes that regulate the host-mediated selection of the microbiome.

Genome-wide association studies (GWAS) represent a powerful approach to map loci that are associated with complex traits in a genetically diverse population. Though pioneered for use in human genetics, to date the majority of GWAS have been conducted in plants (326), and it has become an increasingly popular tool for studying the genetic basis of natural variation and traits of agricultural importance. When inbred lines are available, GWAS can be particularly useful; once genotyped, these lines can be phenotyped multiple times, making it possible to study many different traits in many different environments (327). While GWAS is typically used in the context of a single quantitative phenotypic trait, analyses of multivariate molecular traits, such as transcriptomic or metabolomic data, have also been conducted (328,329). More recently, several attempts have been made to use host-associated microbiome census data as an input to GWAS, which in theory will allow for the identification of host genetic loci controlling microbiome composition (330,331).

In plants, a recent study in Arabidopsis thaliana used phyllosphere microbial community data as the phenotypic trait in a GWAS to demonstrate that plant loci responsible for defense and cell wall integrity affect microbial community variation (332). Several other recent phyllosphere studies performed GWAS to identify genetic factors controlling microbiome associations with mixed degrees of success (332–334). Compared to studies of the phyllosphere, GWAS on Arabidopsis root microbes identified host SNPs associated within and surrounding genes with characterized roles in immunity, cell-wall integrity, and development (335). Previous work comparing the root microbiomes of diverse cereal crops has offered conflicting evidence as to whether host genotypic distance correlates most strongly with microbial community distance within root endospheres or rhizospheres (320,322). These data suggest that the sample type exhibiting the strongest correlation between genotype and microbiome composition may differ for each host and that an initial evaluation of the degree of correlation between genotype and microbiome phenotype across sample types may be informative.

However, to our knowledge, the use of GWAS in conjunction with the rhizosphere microbiome has yet to be explored.

In the context of the root and rhizosphere, we propose Sorghum bicolor (L.) as an ideal plant system for GWAS-based dissection of host-genetic control of microbiome composition. Sorghum is a heavy producer of root exudates (336), and the sorghum microbiome has been shown to house an unusually large number of host-specific microbes (321). In addition, there is a wide range of natural adaptation in traditional sorghum varieties from across Africa and Asia, and a collection of breeding lines generated from U.S. sorghum breeding programs, both of which provide a rich source of phenotypic and genotypic variation (337). Several genome sequences of sorghum varieties have been completed, and variation in nucleotide diversity, linkage disequilibrium (LD), and recombination rates across the genome has been quantified (338), providing an understanding of the genomic patterns of diversification in sorghum. Finally, sorghum is an important cereal crop grown throughout the world as a food, feedstock, and biofuel, enabling direct integration of resulting discoveries into an agriculturally relevant system.

In this study, we dissect the host-genetic control of bacterial microbiome composition in the sorghum rhizosphere. Using 16S rRNA sequencing, we profiled the microbiome of a panel of 200 diverse genotypes of field-grown sorghum. We test the hypothesis that a subset of the sorghum microbiome responds to the host genotype and demonstrate that this subset shares considerable overlap with lineages shown to be susceptible to host genetic control in another plant host. In addition, we tested whether GWAS can be used to identify specific genetic loci within the host genome that are correlated with the abundance of specific heritable lineages and whether differences in microbiome composition can be predicted solely from genotypic information. Collectively, this work demonstrates the utility of GWAS for analyzing host-mediated control of rhizosphere microbiome phenotypes.

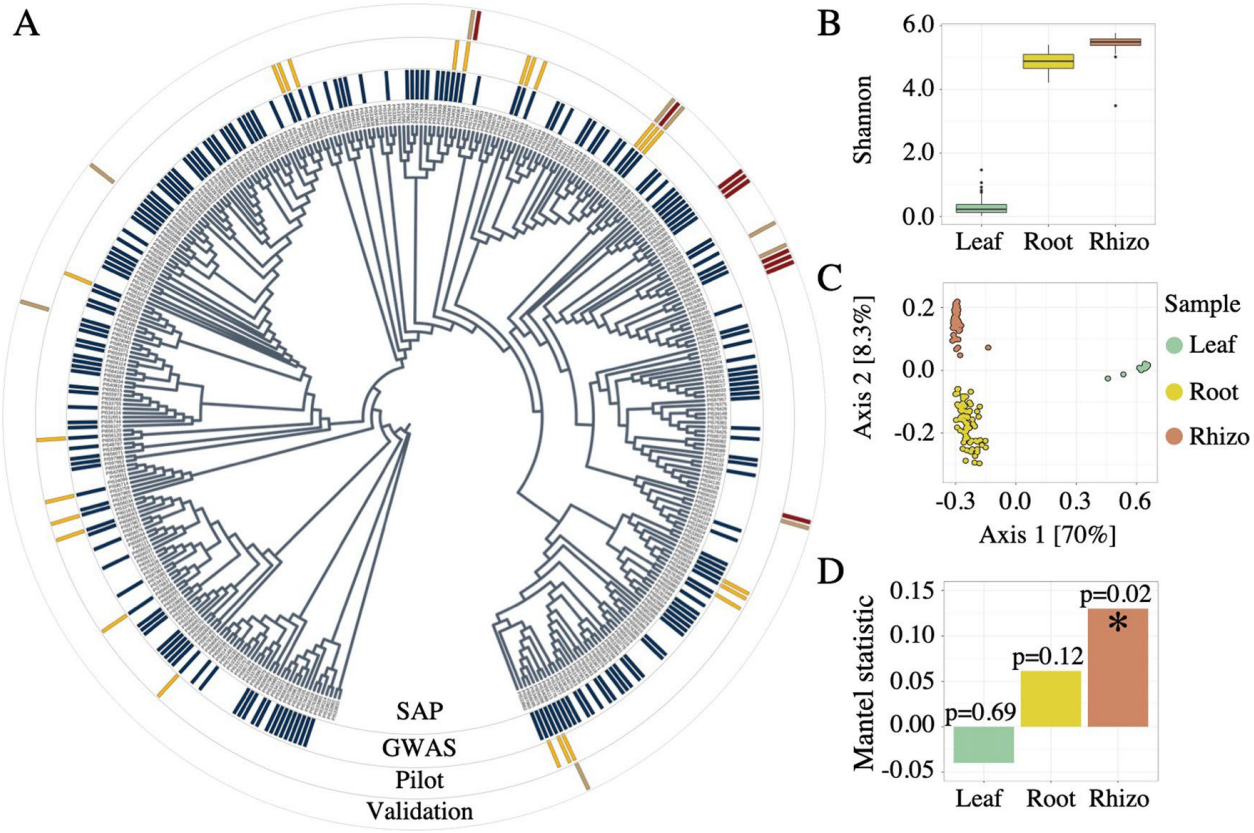
2.3.3 Methods

2.3.3.1 Germplasm selection

In order to ensure that microbiome profiling was performed on a representative subset of the broad genetic diversity present in the 378 members sorghum association panel (SAP) (337,338), subsets of 200 genotypes were randomly sampled from the SAP 10,000 times and an aggregate nucleotide diversity score was calculated for each using the R package “PopGenome” (339). From these data, the subset of 200 lines with the maximum diversity value was selected (Figure 2-3a, Supplementary Figure 2-3, Supplementary Table 2-1). For the pilot experiment that was used to determine the appropriate sample type for GWAS, a subset of 24 lines was selected that included genotypes from a wide range of phylogenetic distances (Figure 2-3a, Supplementary Table 2-1). The phylogenetic tree of sorghum accessions was generated by the neighbor-joining method using an identity by state (IBS) distance matrix calculated in TASSEL 5.0 (340) and visualized using the online tool: interactive tree of life (iTOL) v5 (ref. (341).

2.3.3.1 Sample type and population selection

Figure 2-3. Sample type and population selection.



A Phylogenetic tree representing the 378 member sorghum association panel (SAP, inner ring), the subset of 200 lines selected for GWAS (second ring from the center, in blue), the 24 lines used for sample type selection (Pilot, third ring from the center, in yellow), and the 18 genotypes used for GWAS validation containing either the Chromosome 4 minor allele (red) or major allele (brown) identified by GWAS (outer ring). **B** Shannon's entropy values from 16S rRNA amplicon datasets for the leaf (green), root (yellow), and rhizosphere (red) sample types across all 24 genotypes used in the pilot experiment. **C** Principal coordinate analysis generated using Bray–Curtis distance for the 24 genotypes across leaf (green), root (yellow), and rhizosphere (red). **D** Mantel's R statistic plotted for each sample type across all 24 genotypes indicating the degree of correlation between host genotypic distance and microbiome distance.

2.3.3.2 Field experimental design and root microbiome sample collection

The experimental field used in this study is an agricultural field site located in Albany, California (37.8864 °N, 122.2982 °W), characterized by a silty loam soil with pH 5.2 (321). Germplasm for the US SAP panel used in this study (337) was obtained from GRIN (www.ars-grin.gov). To ensure a uniform starting soil microbiome for all sorghum seedlings and to control their planting density, seeds were first sown into a thoroughly homogenized field soil mix in a growth chamber with controlled environmental factors (25 °C, 16 h photoperiods) followed by transplantation to the field site. To prepare the soil for seed germination, 0.54 cubic meters of soil was collected at a depth of 0–20 cm from the field site subsequently used for planting, and homogenized by separately mixing four equally sized batches with irrigation water in a sterilized cement mixer

followed by manual homogenization on a sterilized tarp surface. The soil was then transferred to sterilized 72-cell plant trays. To prepare seeds for planting, seeds were surface-sterilized through soaking 10 min in 10% bleach +0.1% Tween-20, followed by four washes in sterile water. Following planting, sorghum seedlings were watered with ~5 ml of water using a mist nozzle every 24 h for the first 3 days, and bottom watered every three days until the 12th day, then transplanted to the field.

The field consisted of three replicate blocks, with each block containing 200 plots for each of 200 selected genotypes. Six healthy sorghum seedlings of each genotype were transplanted to their respective plots, separated by 15.2 cm, and thinning to three seedlings per plot was performed at two weeks post transplanting. Plots were organized in an alternating pattern with respect to the irrigation line to maximize the distance between each plant (Supplementary Figure 2-2). Plants were watered for 1 h, three times per week, using drip irrigation with 1.89 L/h rate flow emitters. Manual weeding was performed three times per week throughout the growing season. To ensure that the genotypes were at a similar stage of development and that the host-associated microbiome had sufficient time to develop, collection of plant-associated samples was performed nine weeks post-germination. Only the middle plant within each plot was harvested to help mitigate potential confounding plant-plant interaction effects resulting from contact with roots from neighboring plants of other genotypes. Rhizosphere, leaf, and root samples were collected as described in detail previously (342).

2.3.3.3 DNA extraction, PCR amplification, and Illumina sequencing

DNA extractions, PCR amplification of the V3–V4 region of the 16S rRNA gene, and amplicon pooling were performed as described in detail previously (342). In brief, DNA extractions for all samples were performed using extraction kits (MoBio PowerSoil DNA Isolation Kit, MoBio Inc., Carlsbad, CA) following the manufacturer's protocol. Amplification of the V3–V4 region of the 16S rRNA gene was performed using dual-indexed 16s rRNA Illumina iTags primers 341F (5′–CCTACGGGNBGCASCAG–3′) and 785R (5′–GACTACNVGGGTATCTAATCC–3′). An aliquot of the pooled amplicons was diluted to 10 nM in 30 µL total volume before submitting to the QB3 Vincent J. Coates Genomics Sequencing Laboratory facility at the University of California, Berkeley for sequencing using Illumina Miseq 300 bp pair-end with v3 chemistry. Sequences were returned demultiplexed, with adaptors removed.

2.3.3.4 Amplicon sequence processing, OTU classification, and taxonomic assignment

Sequencing data were analyzed using the iTagger pipeline to obtain OTUs (343). In brief, after filtering 81,416,218 16S rRNA raw reads for known contaminants (Illumina adapter sequence and PhiX), primer sequences were trimmed from the 5′ ends of both forward and reverse reads. Low-quality bases were trimmed from the 3′ ends prior to assembly of forward and reverse reads with FLASH (344). The remaining 66,524,451 high-quality merged reads were clustered with simultaneous chimera removal using UPARSE (345). After clustering, 37,867,921 read counts mapped to operational taxonomic units (OTUs) at 97% identity (Supplementary Table 2-2). Taxonomies were assigned to each OTU using the RDP Naïve Bayesian Classifier with custom reference

databases (346). For the 16S rRNA V3–V4 data, this database was compiled from the May 2013 version of the GreenGenes 16S database v13, trimmed to the V3–V4 region. After taxonomies were assigned to each OTU, OTUs were discarded if they were not assigned a Kingdom level RDP classification score of at least 0.5, or if they were not assigned to Kingdom Bacteria, which yielded 10,006 OTUs. In the downstream analyses, we removed low abundance OTUs (<3 reads in at least 20% of the samples) because in many cases they are artifacts generated through the sequencing process (347). Samples with low read counts were also removed. To adjust for differences in sequencing depth and fit a normal distribution, samples for heritability and GWAS analyses were normalized by cumulative sum scaling (348). For all other analyses, samples were rarefied to an even read depth of 18,000 reads per sample.

2.3.3.5 Estimates of broad-sense heritability of OTU abundance in rhizosphere

To calculate the broad-sense heritability (H2) for individual OTU abundances, we fit the following linear mixed model to OTU abundances of each individual OTU (n = 1189) following a cumulative sum scaling (348) normalization procedure that adjusted for differences in sequencing depth and fit a normal distribution:

$$Y_{ijk} = u + Gi + Bjk + e$$

In this model for a given OTU, Y_{ijk} denotes the OTU abundance of the *i*th genotype evaluated in the *k*th block of the *j*th replicate; *u* denotes the overall mean; Gi is the random effect of the *i*th genotype; Bjk is the random effect of the *j*th replicate nested within the *k*th block; *e* denotes the residual error. With such a model, we divided the environmental variance by the number of replicates according to (349). In order to model the spatial trends in the field, we used the 2-dimensional splines approach that was proposed to accommodate the field's spatial effects by (350,351). This approach was implemented in the R package "sommer" (352) that we used for the model fitting. The variance explained by the spatial effects was excluded from the environmental variance for heritability calculation. To get the null distribution of H2, OTU abundances were randomly shuffled 1000 times and then fitted to the same model as described above. Permutation p-value was calculated as the probability the permuted H2 values were bigger than the observed H2 value.

2.3.3.6 Comparative analysis of heritable taxa between sorghum and maize datasets

To identify the degree to which highly heritable taxa were shared between maize and sorghum, we compared the top 100 most heritable OTUs reported from both maize datasets (referred to as NAM 2010 and NAM 2015) and the sorghum dataset generated in this study. This cutoff, which resulted in a more stringent cutoff than H2 > 0.15, was used for this meta-analysis because H2 varied widely between NAM 2010 and NAM 2015, and a single absolute H2 cutoff would otherwise bias the number of reported highly heritable OTUs between studies. For this analysis, the reported OTUs were aggregated at the order level, resulting in a combined dataset of 300 OTUs spanning 65 bacterial orders. The order level was selected for the following reasons. First, primer differences between studies (V4 in maize compared with V3–V4 in this study), will

impact both phylogenetic assignment and resolution at lower taxonomic ranks (353). Second, taxonomic classification below order was not available for all OTUs. For example, only 14/300 OTUs (4.7%) were not classified beyond phylum or class, while an additional 92 out of the remaining 286 OTUs (32.2%) were not classified below order (Supplementary Table 2-3). Lastly, when orders were separated into lower classifications, lineage membership sizes were insufficient for downstream statistical analyses.

A subset of the orders (n = 18) containing highly heritable OTUs in the maize dataset was not detected in either the high or lowly heritable fractions of the sorghum dataset and was excluded from subsequent comparative analyses. Of the remaining bacterial orders represented by these highly heritable OTUs, we determined the number (n = 26) that contained highly heritable OTUs in at least two of the datasets, and the number (n = 15) that contained highly heritable OTUs in all three datasets. To understand the taxonomic diversity contained within these 65 orders, the family and genus classification of the 300 heritable OTUs are provided in Supplementary Table 2-3. To evaluate whether the degree of overlap in highly heritable lineages is greater than what would be expected by chance, we performed a permutation test (n = 10,000) in which we resampled 100 random OTUs from the 1189 total sorghum OTUs and recomputed intersections with the two maize datasets. These resamplings were not based on OTU abundance; as such, it would be equally likely to draw an abundant or rare OTU, which avoids the possible confounding issue of heritability being correlated with abundance. P-values are reported as the number of instances that these permutations returned a greater degree of overlap in these permutations divided by the total number of permutations.

2.3.3.7 GWAS

A genome-wide SNP map of sorghum SAP accessions used in this study was obtained from the community resource generated previously by genotyping-by-sequencing and included characterization at 265,487 SNPs (338) . GWAS was performed using SNPs with a minor allele frequency (MAF) ≥ 0.01 following (354). For each OTU, GWAS was conducted separately using the best linear unbiased predictors (BLUPs) obtained from the linear mixed model. Population structure was accounted for using statistical methods that allow us to detect both population structure (Q) and relative kinship (K) to control spurious association. The Q model ($y = S\alpha + Qv + e$), the K model ($y = S\alpha + Zu + e$), and the Q + K model ($y = X\beta + S\alpha + Qv + Zu + e$) described previously (355), were used in our study. In the model equations, y is a vector of phenotypic observation; α is a vector of allelic effects; e is a vector of residual effects; v is a vector of population effects; β is a vector of fixed effects other than allelic or population group effects; u is a vector of polygenic background effects; Q is the matrix relating y to v; and X, S, and Z are incidence matrices of 1 s and 0 s relating y to β, α, and u, respectively. To account for the population structure and genetic relatedness, the first three principal components (PCs) and kinship matrix were calculated using the SNPs obtained from (338) and fitted into the MLM-based GWAS pipeline for each OTU using GEMMA (356).

2.3.3.8 GWAS validation experiment

For the GWAS validation experiment, the 378 genotypes of the SAP were the first subset into lines containing the major (n = 343) and minor (n = 14) allele for the two haplotypes found at the peak on chromosome 4 described in the text. Including the 178 genotypes not selected for the GWAS, a total of nine sorghum genotypes belonging to the minor allele were selected, with an effort to include genotypes spanning the phylogenetic tree. For each of these nine minor allele lines, another genotype containing the major allele with close overall genetic relatedness was selected, resulting in nine major and nine minor allele-containing lines. Two replicates of each line were grown in growth chambers (33 °C/28 °C, 16 h light/8 h dark, 60% humidity) in a 10% vermiculite/90% calcined clay mixture rinsed with a soil wash prepared from a 2:1 ratio of field soil to water from the field site used in the GWAS. Plants were watered daily with approximately 5 ml of autoclaved Milli-Q water using a spray bottle for the first 3 days, followed by top watering with 15 ml of water every three days. An additional misting was performed to the soil surface every 24 h to prevent drying. Following two weeks of growth, plants were harvested and rhizosphere microbiomes extracted as described for the field experiment.

2.3.3.9 Microbiome statistical analyses

All statistical analyses of the amplicon datasets were performed in R using the normalized reduced dataset unless stated otherwise. For alpha-diversity measurement, Shannon's Diversity was calculated with the diversity function in the R package vegan (357). Principal coordinate analyses were performed with the function pcoa in the R package ape (358), using the Bray–Curtis distance obtained from the function vegdist in the R package vegan (357). Mantel's tests were used to determine the correlation between host phylogenetic distances and microbiome distances using the mantel function in the R package vegan (357) with 9999 permutations, and using Spearman's correlations to reduce the effect of outliers. Indicator species analyses were performed using the function indval in the R package labdsv(359), with p-values based on permutation tests run with 10,000 permutations. Multiple testing corrections were performed with an FDR of 0.05 using the p.adjust function in the base R package stats. Canonical analysis of principal coordinates (CAP) was performed using the capscale function in the R package vegan (357); an ANOVA like permutation test using the sum of all constrained eigenvalues was performed to determine the percent variance explained by each factor using the function anova.cca in the R package vegan (357) .

2.3.3.10 Analysis of sorghum RNA-seq datasets

Publicly available sorghum RNA-Seq data for 27 annotated genes in the 1.15 Mb interval of chromosome 4 (Sobic.004G153000–Sobic.004G155900), were downloaded from phytozome v12.1 (ref. (360)). Expression datasets were broadly grouped based on the tissue type from which they were derived (root, leaf, or reproductive). To aid in the visualization of tissue-specific expression of genes exhibiting large differences in absolute levels of gene expression, we normalized the fragments per kilobase of transcript per million mapped reads (FPKM) values for each gene in each tissue type by dividing by the average value of gene expression for that gene across all tissue types. We defined root-specific expression as genes that had a normalized FPKM less than 1

in no more than two root datasets, and a normalized FPKM greater than 1 in no more than two datasets of other tissue types.

2.3.4 Results

2.3.4.1 Diverse sorghum germplasm show rhizosphere is ideal for microbiome-based GWAS

In this study, the relationship between host genotype and microbiome composition was explored through a field experiment involving 200 genotypes selected from the SAP germplasm collection (337,338) (Supplementary Table 2-1). While a recent study in Arabidopsis successfully performed GWAS using the root microbiome (endosphere) (335), it did not evaluate microbes that are closely associated with the exterior of the root (rhizosphere). We first sought to determine whether leaf, root endosphere, or rhizosphere samples were most suitable for downstream GWAS in sorghum. Using a subset of 24 genotypes from our collection of 200 (Figure 2-3a, Supplementary Table 2-1), the microbiome composition of leaf, root, and rhizosphere sample types were analyzed using paired-end sequencing of the V3–V4 region of the ribosomal 16S rRNA. The resulting dataset demonstrated comparatively high levels of microbial diversity within both root and rhizosphere samples (Figure 2-3b) and strong clustering of above and below ground sample types (Figure 2-3c). Three independent Mantel's tests (9999 permutations) were used to evaluate the degree of correlation between host genotypic distance and microbiome composition for leaf, root, and rhizosphere sample types (Figure 2-3d); of the three compartments, the only rhizosphere exhibited a significant Mantel's correlation ($R2 = 0.13$, $Df = 1$, $p = 0.02$). Based on these results, subsequent investigation of the microbiomes of the full panel of 200 lines, including heritability and GWAS analyses, was performed using rhizosphere samples.

To investigate host genotype-dependent variation in the sorghum rhizosphere microbiome, the rhizospheres of 600 field-grown plants (three replicates of each of the 200 genotypes) were profiled using V3–V4 16S rRNA amplicon sequencing. The resulting data set included 1189 OTUs representing 29 bacterial phyla. Compositional analysis of the microbiome dataset exhibited profiles consistent with recent microbiome studies involving the sorghum rhizosphere (321,361,362) from a variety of field sites, with proteobacteria, actinobacteria, and acidobacteria comprising the top three dominant phyla (Supplementary Figure 2-3).

2.3.4.2 Sorghum and maize rhizospheres exhibit strong overlap in heritable taxa

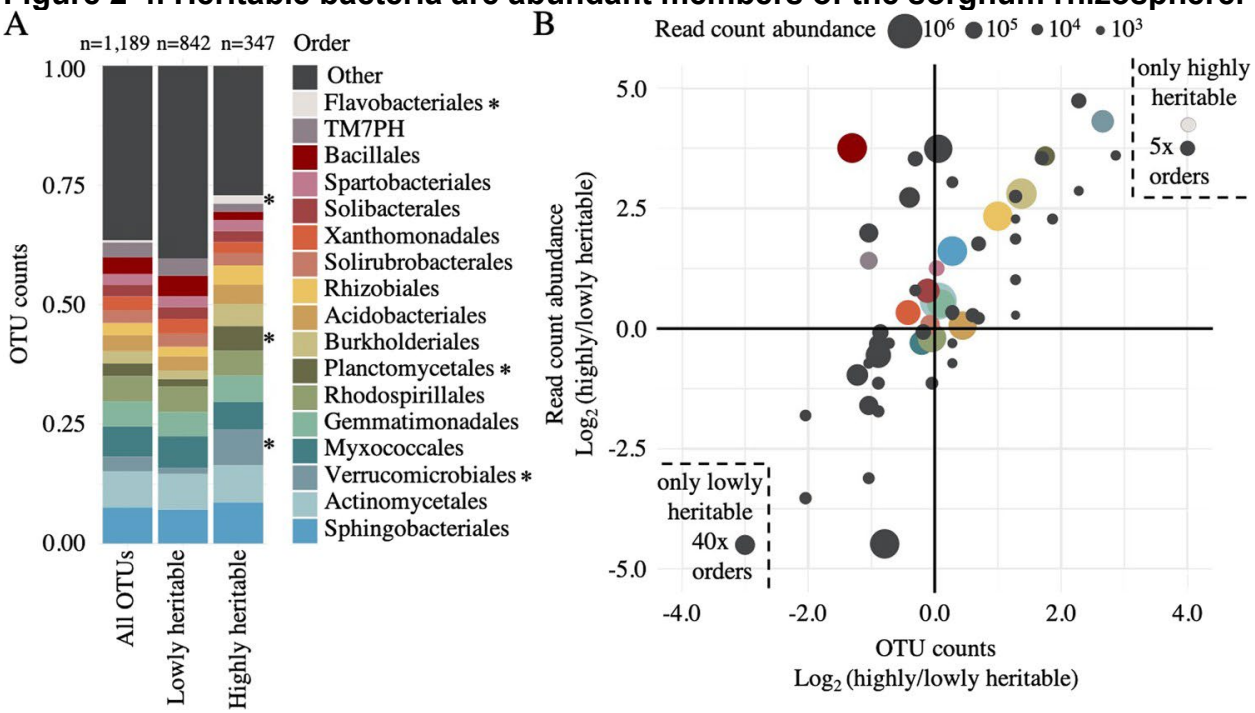
A recent study of two separate maize microbiome datasets suggests that specific bacterial lineages are more sensitive to the effect of host genotypes than others (322). To determine if a bacterial lineage's responsiveness to host genetics is a trait conserved across different plant hosts, the broad-sense heritability (H2) of individual OTUs in our sorghum dataset was evaluated; H2 quantifies the proportion of variance that is explained by genetic rather than environmental effects. More generally, this approach treats bacterial abundance as a continuously varying phenotype, similar to plant height, biomass, and yield (363). In our study, H2 ranged from 0 to 66% for individual OTUs (Supplementary Table 2-4). By comparison, H2 for individual OTUs in the first of two experiments across 27 maize inbred lines had a maximum of 23% (performed in 2010),

while the second exhibited a maximum of 54% (performed in 2015) (322). Further, we used the sorghum diversity panel kinship matrix to calculate the SNP-based narrow-sense heritability (h2). Consistent with our expectation, the h2 was lower than the H2 estimated from the phenotypic data, likely due to overcorrection of the spatial components in the spline analysis. Nevertheless, h2 was significantly ($R = 0.34$, $p = 2.2e{-}16$) correlated with previous phenotype-based estimates (Supplementary Figure 2-4a), especially for the top 100 heritable OTUs ($R = 0.5$, $p = 1e{-}7$, Supplementary Figure 2-4b).

To explore whether microbes with high heritability in the sorghum dataset are phylogenetically clustered, we partitioned the 1189 OTUs into highly heritable ($n = 347$) and lowly heritable fractions ($n = 842$) using an H2 cutoff score of 0.15 (Figure 2-4a, Supplementary Table 2-5). Several bacterial orders, including verrucomicrobiales, flavobacteriales, and planctomycetales, were observed to have significantly greater numbers of OTUs that are highly heritable, as compared to the lowly heritable OTU fraction (Fisher's exact test, $q < 0.05$, Figure 2-4a, Supplementary Table 2-5). Notably, all 6 Flavobacteriales OTUs were only present in the highly heritable fraction (Figure 2-4b); by contrast, 40 other bacterial orders were only observed within the lowly heritable fraction. Having established that some bacterial lineages had a higher proportion of OTUs that were highly heritable, we aimed to determine what fraction of the total read count abundance these heritable OTUs represented. In general, we observed that read count abundance per taxa correlated with heritability, with some exceptions (Figure 2-4b). For example, Bacillalles, contained a smaller number of OTUs in the highly heritable than a lowly heritable fraction, but the percentage of reading counts attributable to its highly heritable OTUs was approximately eight-fold greater than those in the lowly heritable fraction, suggesting that its highly heritable members are abundant organisms within the rhizosphere (Figure 2-4b). Collectively, these data imply that a specific subset of bacterial lineages is enriched for members susceptible to host genotypic selection.

2.3.4.3 Heritable bacteria are abundant members of the sorghum rhizosphere

Figure 2-4. Heritable bacteria are abundant members of the sorghum rhizosphere.



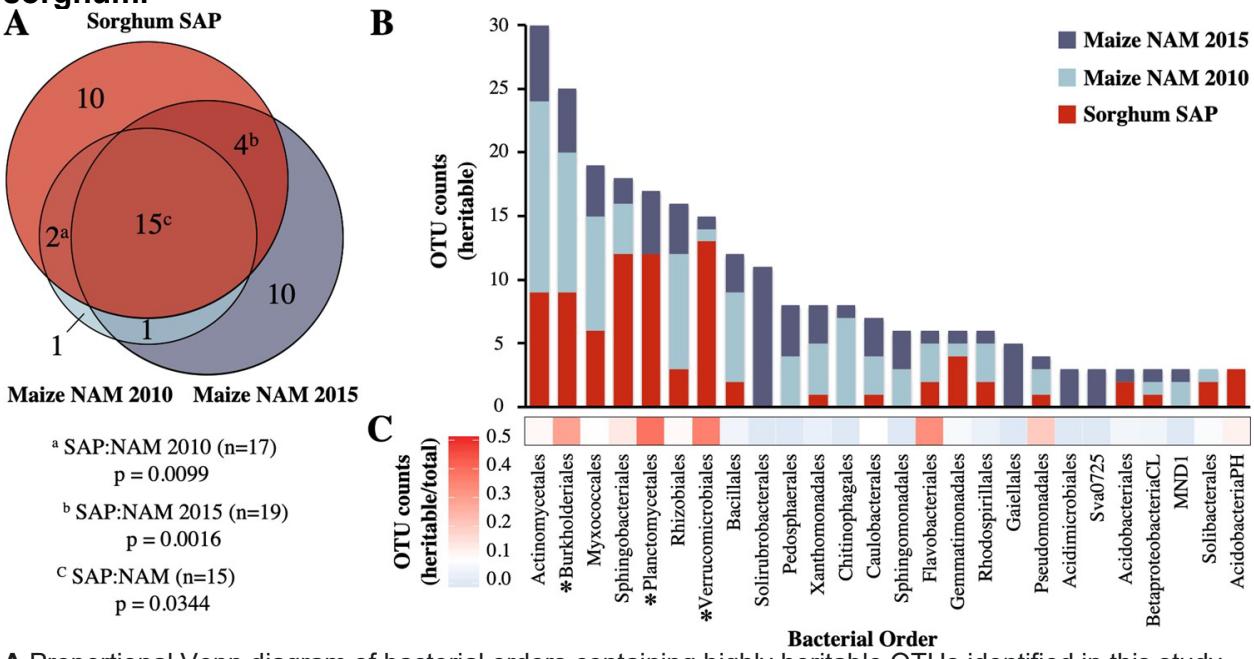
A The relative percentage of total OTUs belonging to each of the top 17 bacterial orders for all OTUs (left bar), lowly heritable OTUs (middle bar), or highly heritable OTUs (right bar). Orders with significantly different numbers of OTUs in the highly heritable (H2 > 0.15) as compared to the lowly heritable fraction (H2 < 0.15), as determined by Fisher's exact test (q < 0.05), are indicated with asterisks. **B** Order-level scatterplot of the log2 ratio between highly and lowly heritable OTU counts (x-axis) and read count abundance (y-axis). Circle sizes represent the total read count abundance of each bacterial order. Bacterial taxa that were present only in the highly heritable (upper right, Flavobacteriales and 5 "other" merged orders) or lowly heritable (lower left, 40 "other" merged orders) fractions of the dataset are presented within the dashed lines, as their Log2 heritability ratios are undefined numbers.

Next, we hypothesized that despite the considerable evolutionary distance between maize and sorghum (two members of the grass family Poaceae that diverged more than 11 million years ago (364)), the bacterial lineages containing OTUs most responsive to host genotypic effects in maize would likely also contain OTUs exhibiting such susceptibility within sorghum. To test this, we compared the top 100 most heritable OTUs from both maize datasets (referred to as NAM 2010 and NAM 2015) and the sorghum dataset described above, resulting in a combined dataset of 300 OTUs spanning 65 bacterial orders. After removing bacterial orders not observed in the sorghum dataset (n = 18), we noted that more than half were observed in at least two of the datasets, and approximately one-third (n = 15) contained highly heritable OTUs in all three datasets (Figure 2-5a). To determine if this overlap was significantly greater than is expected by chance, we performed permutational resampling of 10,000 sets of randomly chosen sorghum OTUs for comparison. Notably, we found that the overlap between the highly heritable sorghum fraction with both the individual maize heritable fractions and the combined heritable maize OTUs to be significant, compared with the resampled sorghum OTUs (NAM 2010 n = 17, p = 0.0099, NAM 2015 n = 19, p = 0.0016, combined n = 15, p = 0.0344) (Figure 2-5a). Collectively, these results imply that there is

conservation between the bacterial orders most sensitive to genotype across both maize and sorghum.

2.3.4.4 Heritability of rhizosphere microbes is conserved across maize and sorghum

Figure 2-5. Heritability of rhizosphere microbes is conserved across maize and sorghum.



A Proportional Venn diagram of bacterial orders containing highly heritable OTUs identified in this study (Sorghum SAP), compared with the heritable orders reported in a large-scale field study of maize nested association mapping (NAM) parental lines grown over two separate years, published in Walters et al.(322). The top 100 heritable OTUs (based on H2) from each dataset were classified at the taxonomic rank of order to generate the Venn diagram. NAM highly heritable orders only present in the SAP lowly heritable fraction are represented by the blue sections. Superscript letters indicate the frequency that a random subsampling of 100 sorghum OTUs from the total 1189 sorghum OTUs (10,000 permutations) produced greater order-level overlap with maize OTUs from either single year (a/b) or both (c). **B** Stacked barplot displaying cumulative counts (y-axis) of OTUs identified as highly heritable in any of the three datasets for all bacterial orders (x-axis) which have a total of at least three highly heritable OTUs in an order. **C** The fraction of highly heritable sorghum OTUs relative to all sorghum OTUs within each order is displayed as a heatmap. Asterisks indicate orders enriched in highly heritable OTUs (Fisher's exact test, $q < 0.05$).

In an effort to identify the bacterial lineages with the greatest propensity for high heritability, we calculated the number of highly heritable OTUs in each of the shared highly heritable bacterial orders identified above. We noted that among bacterial orders containing the greatest number of highly heritable OTUs across all three datasets were several that represent large lineages frequently observed within the root microbiome; (e.g., actinomycetales) (Figure 2-5b). We hypothesized that this result is likely driven in part by the overall frequency of these lineages within the rhizosphere microbiome, with more common lineages resulting in a greater fraction of highly heritable microbes due to their ubiquity. To help account for this, we normalized the frequency of highly heritable sorghum OTUs (n = 100) by total sorghum OTU counts (n = 1189) belonging to each

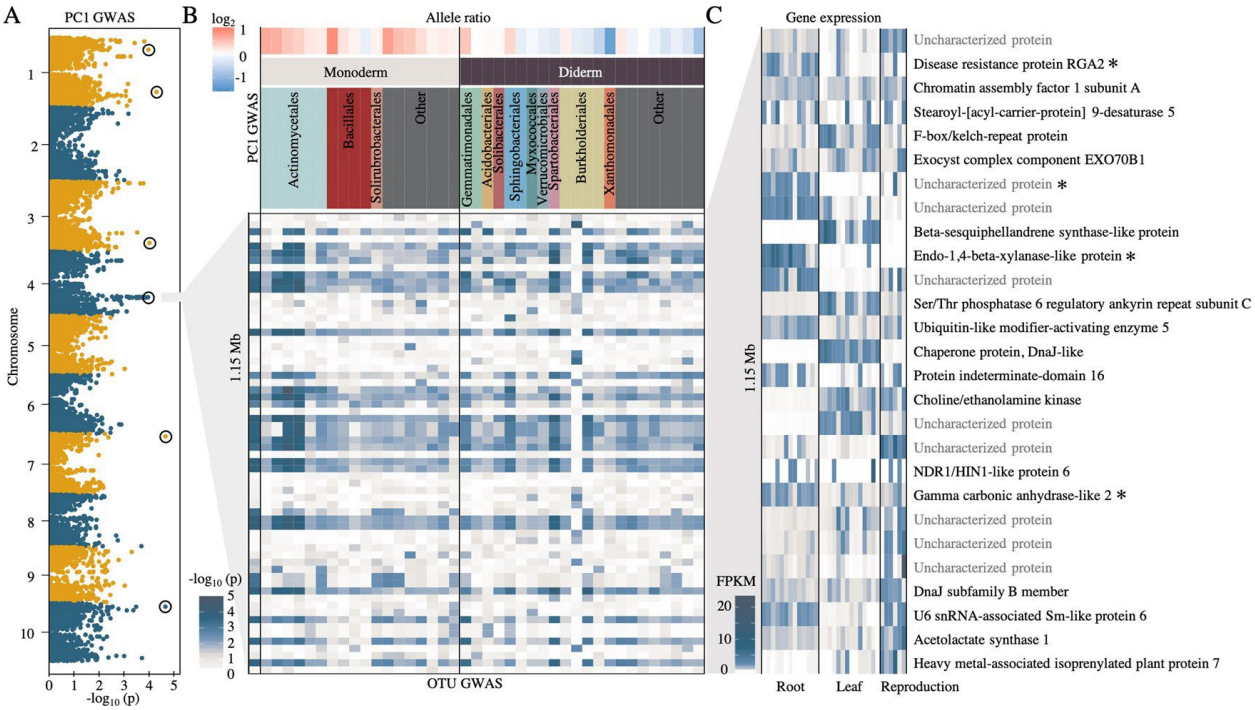
order (Figure 2-5c, Supplementary Table 2-6). These results demonstrate that while the prevalence of actinomycetales and myxococcales among highly heritable microbes is consistent with their general prevalence in the overall dataset, Burkholderiales and two other lineages, including the verrucomicrobia and planctomycetes, exhibited a significant enrichment (Fisher's exact test, $q < 0.05$) in the highly heritable fraction not expected to be influenced by abundance alone.

2.3.4.5 Genome-wide association reveals genetic loci correlated with rhizosphere microbial abundance

Recent work in the leaf microbiome has demonstrated the potential utility of GWAS for uncovering host loci correlated with microbiome composition (334). Here, we sought to use GWAS with rhizosphere microbiome datasets using both global properties of the OTU dataset and the abundances of individual OTUs. For overall community composition, a subset of PCs was selected from an analysis of the abundance patterns of the 1189 OTUs. To prioritize individual PCs for inclusion in our GWAS analysis, we determined the heritability scores of each of the top ten PCs, which explained 75% of the total variance in our dataset (Supplementary Figure 2-7a). PCs with H2 equal to or greater than 0.25 (PC1, PC3, PC5, PC9, and PC10, Supplementary Figure 2-7a) were subjected to GWAS (Supplementary Figure 2-7b). We initially applied a strict Bonferroni-based threshold (adjusted p-value $< 0.05/21{,}236$) to our GWAS. However, no significant SNPs were identified using this method, suggesting this threshold was too stringent and masking potential true positive associations, due to a combination of small sample size ($n = 200$) and relatively low heritability ($H2 = 0.35$ for PC1) in our data. Despite the common use of applying multiple testing corrections, including Bonferroni corrections, to GWAS to define significance cutoffs, it is understood that these cutoffs are overly conservative due to the assumption that every genetic variant tested is independent of the rest (365,366). To discover potential true associations that were missed by Bonferroni correction, we applied an anti-conservative false discovery rate cutoff of $-\log10\ (p = 10^{-4})$ (Benjamini–Hochberg, $q < 0.22$) to generate a list of top candidate SNPs. The GWAS analysis performed for PC1, which explained 21% of the total variance and had the second-highest heritability ($H2 = 0.35$), revealed a correlation between community composition and a locus on chromosome 4 that was among these top candidates ($n = 6$) (Figure 2-6a, Supplementary Figure 2-6a).

2.3.4.6 A sorghum genetic locus is correlated with rhizosphere microbial abundance

Figure 2-6. A sorghum genetic locus is correlated with rhizosphere microbial abundance.



A Manhattan plot of PC1 community analysis GWAS. Top candidate SNPs above a threshold of –log10 (p = 10−4) are circled. **B** Individual OTU GWAS of all OTUs with at least 5 SNPs above a threshold of –log10 (p = 10−2.5) in the 1.15 Mb window identified on the same chromosome 4 locus identified by PC1 GWAS (lower heatmap). For each OTU, the log2 fold change in abundance between the sorghum major (red) or minor (blue) allele groups within this locus was determined (upper heat map). OTUs were grouped based on the predicted presence of one or two membranes (monoderm or diderm) within each bacterial order and colored as in Figure 2. **C** Tissue-specific gene expression data for sorghum genes within the chromosome 4 locus. Darker blue indicates higher expression (normalized FPKM). Asterisks indicate genes whose expression is predicted to be root-specific.

The MAF of candidate SNPs at the chromosome 4 locus ranged from 0.021 to 0.036. We carried out LD analysis between the most significant SNP (leading SNP) and the SNPs around the region. Results revealed that SNPs exhibiting high LD with the leading SNP were physically close (<1.6 Mb) and showed relatively high association signals (Supplementary Figure 2-6b). As SNP imputation leverages local LD information, a cluster of SNPs in LD may be caused by SNP imputation. However, SNP imputation tends to be poor for minor alleles because less data is available to impute compared with common alleles (367). Therefore, the presence of multiple GWAS signals exhibiting different allele frequencies at the chromosome 4 locus suggests the leading SNP was less likely to be a statistical artifact. Using the chromosome 4 locus SNP data, we separated sorghum genotypes into two allele groups, the major allele containing 343 sorghum genotypes and the minor allele containing 14 genotypes, with six minor allele-containing lines present in our 200 line GWAS subset. Using these six minor allele-containing genotypes and a closely related major allele genotype for each minor allele genotype, we performed a CAP ordination of the rhizosphere microbiome, constrained by allele group. This separated the rhizospheres of genotypes belonging to major and minor allele groups into distinct clusters (Supplementary Figure 2-6c).

Subsequent GWAS analyses were performed using the other heritable PCs, PC3, PC5, PC9, and PC10, as inputs (Supplementary Figure 2-7). We did not observe any SNPs below the $q < 0.22$ threshold for these PCs. This low signal is likely in part because the microbiome community has a relatively low heritability, the traits are highly polygenic, and the genome scans involve a large number of statistical tests. However, there was an identifiable peak on chromosome 6 of PC5 (p-value<7 × 10–4) and PC10 (p-value<1 × 10–4) (Supplementary Figure 2-7b). As PCs are derived from linear combinations of the abundance of individual OTUs within the dataset, it is unclear whether the correlations observed on chromosomes 4 and 6 are driven by one common or two different sets of microbial lineages. To address this, we performed separate GWAS analyses using the abundances of each single OTU in our dataset as input (Figure 2-6b, Supplementary Figure 2-7c). From these analyses, we identified two distinct sets of 39 and 10 OTUs with significant correlations with the loci on chromosomes 4 and 6, respectively, and only a single OTU belonging to the order Burkholderiales that was shared between the two loci (Supplementary Figure 2-7c). This implies that different sorghum loci are associated with the abundance patterns of different groups of microbes.

To understand the relationship between the identified peak on chromosome 4 (Figure 2-6a) and the bacterial taxa with similar GWAS correlations at this locus (Figure 2-6b), we first sought to understand how relative abundance for these 40 OTUs varied across the sorghum panel. We observed that the majority of OTUs that were more prevalent in sorghum genotypes containing the major allele belonged to monoderm lineages, while the majority of OTUs more prevalent in the minor allele group belonged to diderm lineages (Figure 2-6b), suggesting that host genetic mechanisms at this locus are interacting with basal bacterial traits.

To explore which host genetic mechanisms might be driving the correlations observed on Chromosome 4, we examined tissue-specific expression patterns from publicly available RNA-Seq datasets obtained from phytozome v12.1 (ref. (360)) for all 27 genes in the 1.15 Mb interval (Figure 2-6c, Supplementary Table 2-7). Of these candidates, we observed compartment-specific expression patterns, including several annotated candidates exhibiting strong root-specific activity: gamma carbonic anhydrase-like 2, a putative beta-1,4 endoxylanase, and disease resistance protein RGA2 (Figure 2-6c).

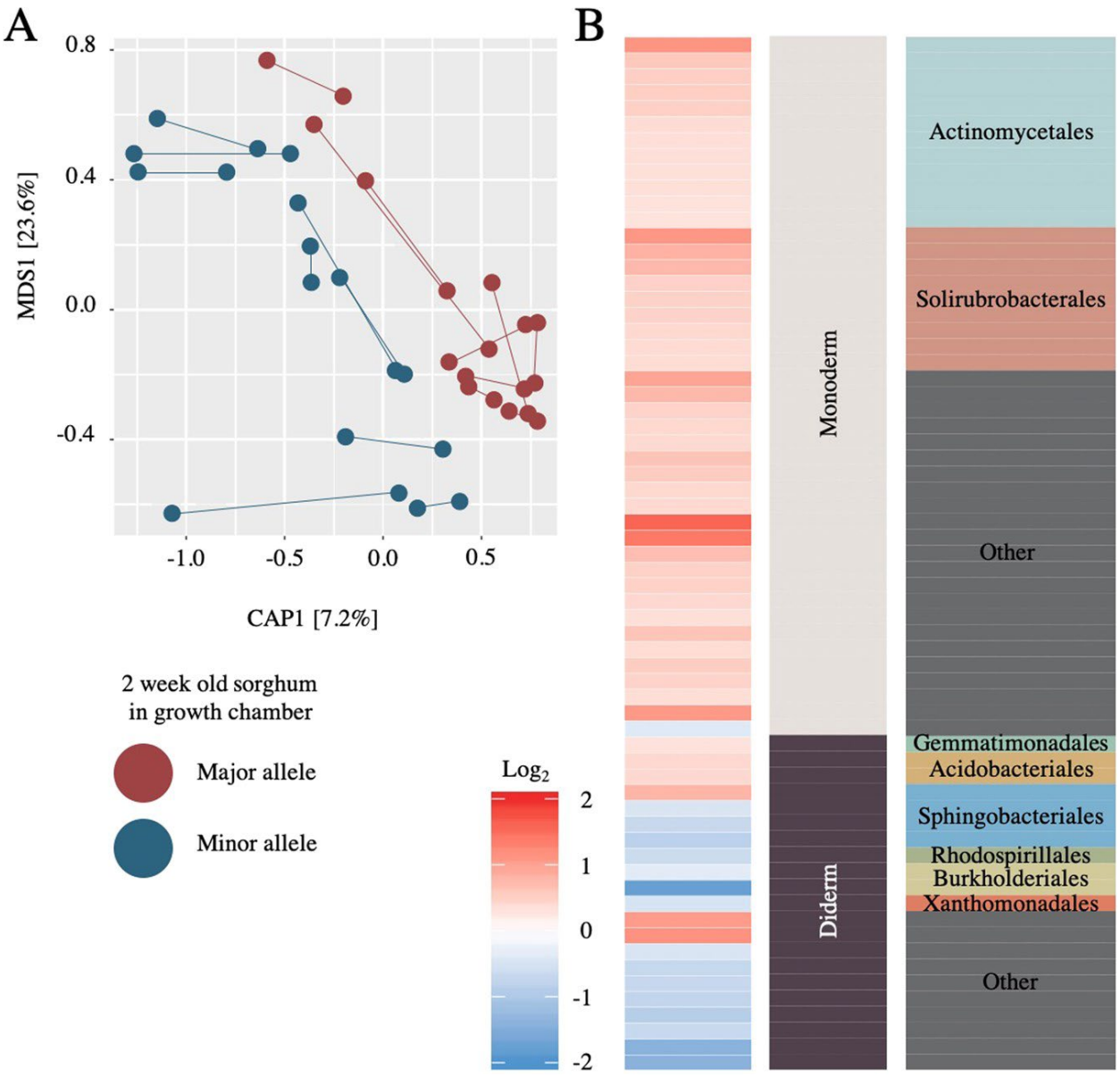
2.3.4.7 Sorghum genotypic data can predict microbiome composition

To validate that allelic variation at the candidate locus on chromosome 4 contributes to differences in rhizosphere composition, we conducted a follow-up growth chamber experiment with eighteen additional sorghum lines, including genotypes not present in the original study. To help disentangle phylogenetic-relatedness from locus-specific effects, we selected sorghum genotypes that spanned the diversity panel; additionally, for each minor allele genotype ($n = 9$), we included a phylogenetically related major allele line ($n = 9$) (Figure 2-3a). Following two weeks of growth in a mixture of calcined clay and field soil in the growth chamber, we collected the rhizosphere microbiomes of each genotype and the microbiome composition was analyzed using 16S rRNA amplicon sequencing as in the main study. We conducted a regional association

analysis using a mixed linear model that included a kinship matrix as the random effect. Using a stringent Bonferroni method, we detected several signals above the threshold that was in high LD with the original GWAS signal (Supplementary Figure 2-8). A CAP ordination constrained on genotypic group separated the rhizospheres of genotypes belonging to major and minor allele groups into distinct clusters (Figure 2-7a, PERMAnova F = 2.66, Df = 1, p = 0.0061), with genotype explaining approximately 7.5% (CAP1) of variance in the dataset.

2.3.4.8 Sorghum genetic information can be used to predict rhizosphere microbiome composition under different growth conditions

Figure 2-7. Sorghum genetic information can be used to predict rhizosphere microbiome composition under different growth conditions.



A Canonical analysis of principal coordinates of the rhizosphere microbiome for nine major allele genotypes (red) and nine minor allele genotypes (blue) grown in a growth chamber for 2 weeks. Two replicates per genotype were used and are connected by lines. **B** For each indicator OTU, the log2 fold change in abundance between the sorghum major (red) or minor (blue) allele groups was determined. OTUs were grouped based on the predicted presence of one or two membranes (monoderm or diderm), within each bacterial order, and colored as in Figs. 2-4 and 2-6.

To identify which taxa drive the clustering observed in our CAP analysis, and to compare this to taxa responsive to the chromosome 4 allele group in our main experiment, we performed an indicator species analysis on the validation dataset. A comparison of the significant indicator OTUs ($q < 0.05$) from each allele group in the validation dataset ($n = 65$) demonstrated similar trends in abundance of indicator OTUs as observed in the main experiment (Figure 2-6b), with OTUs belonging to monoderm and diderm lineages enriched in the major and minor allele-containing lines, respectively. Interestingly, while most diderm lineages were more prevalent in the minor allele-containing lines, several diderm lineages including gemmatimonadales, acidobacteriales, and sphingobacteriales contained OTUs that were more abundant within major allele lines. Notably, this pattern was observed in the rhizospheres of 9-week old field-grown sorghum during the main experiment (Figure 2-6b) and was also observed in 2 weeks growth chamber sorghum in the validation experiment (Figure 2-7b). Collectively, this experiment supports the findings of our main experiment, in which allelic variation at a locus located on chromosome 4 was shown to correlate with the abundance of specific bacterial lineages.

2.3.5 Discussion

2.3.5.1 Host selection of plant rhizosphere microbiomes

Previous GWAS of plant-associated microbiome traits have often been conducted with leaf samples (332–334), and a recent GWAS was applied to root endophytes (335). However, to our knowledge GWAS of the rhizosphere has not been attempted. In this study, we compared the overall correlation between host genotype and bacterial microbiome distances across leaf, root, and rhizosphere of Sorghum bicolor, and demonstrate that of the three, the rhizosphere represents the most promising compartment for conducting experiments to untangle the heritability of the sorghum microbiome. Notably, the degree of correlation between sorghum phylogenetic distance and microbiome distance was highest in the rhizosphere and lowest in the leaves. This greater correlation observed in the root and rhizosphere could be in part due to the phyllosphere's relative compositional simplicity. Even Arabidopsis rosette leaves, which are in close proximity to the soil, harbor a distinct and relatively simple bacterial community compared to the root (335).

By contrast, the rhizosphere represents a highly diverse and populated subset of the soil microbiome and potentially offers a greater pool of microbes upon which the host may exert influence (368). Alternatively, the rhizosphere's greater correlation with microbiome composition could be caused by the plant's relatively weaker ability to select epiphytes in its aboveground microbiome; while the arrival of phyllosphere colonists is largely thought to be driven by wind and rainfall dispersal (369), root exudation is known to control chemotaxis and other colonization activities of select

members of the surrounding soil environment. This provides an additional mechanism for host selection of its microbial inhabitants prior to direct interaction with the plant surface (325,370,371). Once in the root, microbes are subjected to additional selective pressures, including evading host immune systems (335,372), that would determine if they persist as endophytes. While we observed the highest correlation sorghum phylogenetic distance and microbiome distance occurred in the rhizosphere, it is possible that other plant hosts may demonstrate the greatest selective influence within tissues other than the rhizosphere. Future efforts to investigate host control of the microbiome through GWAS or related techniques would benefit from a careful selection of sample types following pilot studies designed to explore heritability across different host tissues.

2.3.5.2 Heritable rhizosphere microbes are phylogenetically clustered and similar across hosts

Within the rhizosphere, we demonstrate that microbiome constituents vary in H2, and highly heritable taxa show strong overlap with highly heritable lineages identified in maize, spanning fifteen different bacterial orders (322). In particular, three of these orders, verrucomicrobiales, burkholderiales, and planctomycetales were significantly enriched in the highly heritable fraction of our dataset. As members of burkholderiales can form symbioses with both plant and animal hosts (373,374), and some colonize specific members of a host genus or species (375), it is feasible that such strong relationships necessitated additional genetic discrimination between hosts. Within Burkholderia spp., this could be facilitated by their relatively large pan-genome, with diversity driven by large multi-replicon genomes and abundant genomic islands (376).

These observations suggest that evaluating bacterial heritability may identify new lineages for which close or symbiotic but previously undetected associations with plant hosts exist. For example, we observed several lineages with high heritability that are common in soil, yet prior evidence of plant-microbe interactions in the literature is lacking, including verrucomicrobiales and planctomycetales. Interestingly, high heritability in these lineages might be facilitated by the presence of a recently discovered shared bacterial microcompartment gene cluster present in both Planctomycetes and Verrucomicrobia, which confers the ability to degrade certain plant polysaccharides (377). Indeed, microbiome composition is known to be driven in part by variations in polysaccharide-containing sources including plant cell wall components and root exudates (378). Additional experimentation with bacterial mutants lacking this genetic cluster could be useful for revealing its role in shaping plant microbe interactions. Finally, we note that these results were generated in part through comparisons of datasets generated from two independent studies with different experimental designs and analysis pipelines; we anticipate that future experiments using a common garden approach could improve upon our efforts here to identify common heritable taxa across plant host lineages.

2.3.5.3 Sorghum loci are responsible for controlling the rhizobiome

Our GWAS correlated host genetic loci and the abundance of specific bacteria within the host microbiome, as well as overall rhizosphere community structure. Our study

builds upon previous research that applied GWAS to the Arabidopsis root microbiome (335), demonstrating that the use of this technique can be further expanded to the rhizosphere of a cereal crop plant. Using this method, we identified a locus on chromosome 4 that was correlated with specific bacterial lineages. Notably, we detected a similar association in a cross-validation experiment, which included both independent genotypes and different environmental conditions, providing strong support that this locus was a true positive, despite the modest statistical stringency applied to GWAS in this study. We observed several annotated candidate genes within the chromosome 4 locus exhibiting strong root-specific activity including gamma carbonic anhydrase-like 2, a putative beta-1,4 endoxylanase, and disease resistance protein RGA2. However, inferences of causal genes based on gene expression patterns come with significant limitations, as there is no requirement that a gene controlling this association would solely be expressed in roots. For example, architectural or hormonal changes in the plant phyllosphere could drive feed-forward effects on root exudate compositions (286,323). While our cross-validation experiment focused on a single locus on chromosome 4, plants are capable of influencing their microbiomes using a multitude of strategies, and many of these traits are predicted to be complex (i.e., controlled by multiple or many genes) (379). As such, it is notable that we detected five additional candidate loci in the PC1 GWAS alone. In addition, a candidate locus on chromosome 6 was identified in both PC5 and PC10 community analysis GWAS. Strikingly, this locus was associated with a distinct set of microbes from the chromosome 4 locus identified by PC1 GWAS, suggesting that sorghum plants are able to use distinct mechanisms to modulate different groups of microbes. Future validation experiments using genetic mutants within these and other candidate genes can be used to help elucidate the underlying genetic element(s) responsible for the modulation of the rhizosphere microbiome.

2.3.6 Conclusion

Although the underlying host genetic causes of shifts in the microbiome are not well understood, candidate-driven approaches have implicated disease resistance (323,324), nutrient status (324,380,381), sugar signaling (382), and plant age (383,384) as major factors. Non-candidate approaches to link host genetics and microbiome composition, such as GWAS, have the potential to discover novel mechanisms that can be added to this list. Here we show that GWAS can predict rhizosphere microbiome structure based on host genetic information, building on previous studies that have observed inter- and intra-species variation in microbiomes (318,321,322,332,335,378,385–387). Collectively, our study adds to a growing list of evidence that genetic variation within plant host genomes modulates their associated microbiome. We anticipate that GWAS of plant microbiome association will promote a comprehensive understanding of the host molecular mechanisms underlying the assembly of microbiomes and facilitate breeding efforts to promote beneficial microbiomes and improve plant yield.

2.3.7 Acknowledgements

We thank doctors Sam Leiboff, Ling Xu, Edi Wipf, and Tuesday Simmons for their helpful discussions and critical readings of the manuscript. This research was funded by a grant from the US Department of Agriculture (2030-12210-002-00D).

2.3.8 Author Contributions

SD conceived and designed the experiments, performed the experiments, analyzed the data, and prepared figures and/or tables; DC conceived and designed the experiments, analyzed the data, and prepared figures and/or tables; GX analyzed the data and prepared figures and/or tables; LD performed the experiments; LW performed the experiments and analyzed the data; JY conceived and designed the experiments, and analyzed the data; DC-D conceived and designed the experiments, analyzed the data, and prepared figures and/or tables; All authors authored or reviewed drafts of the paper and approved the final draft.

2.3.9 Data availability

All datasets and scripts for analysis are available through github (https://github.com/colemanderr-lab/Deng-2020) and all short-read data has been submitted to the NCBI and can be accessed through BioProject PRJNA612320.

2.3.10 Guide to Supplementary Data Online

Supplementary Figure 2-1. Sorghum phylogenetic tree.
Supplementary Figure 2-2. Field experimental design.
Supplementary Figure 2-3. Composition of the sorghum microbiome across sample types.
Supplementary Figure 2-4. Correlations between SNP-based h2 and phenotype-based H2.
Supplementary Figure 2-5. The abundance of distinct sets of microbes are associated with different sorghum genetic loci.
Supplementary Figure 2-6. Analysis of the chromosome 4 locus identified by PC1 GWAS.
Supplementary Figure 2-7. GWAS analysis of all heritable PCs.
Supplementary Figure 2-8. Regional association analysis results for the first principle component of the microbiome community diversity.
Supplementary Table 2-1. Sorghum Association Panel (SAP) germplasm information of lines used in this study.
Supplementary Table 2-2. 16S rRNA gene read count statistics. Raw reads were processed using iTagger pipeline to obtain OTUs. The number of total reads in the dataset remaining after each analysis step in this pipeline, as well as the percentage of the initial reads that these values represent, are displayed.
Supplementary Table 2-3. Taxonomy of heritable OTUs used to generate figure 3.
Supplementary Table 2-4. Broad sense heritability (H2) of individual OTUs in the sorghum dataset.
Supplementary Table 2-5. Taxonomic classification of rhizosphere microbes used to generate figure 2.
Supplementary Table 2-6. Identification of enriched heritable bacterial orders for figure 3.
Supplementary Table 2-7. Sorghum chromosome 4 locus gene IDs and phytozone gene expression values used to generate figure 4C.

2.4 A Suite of Constitutive Promoters for Tuning Gene Expression in Plants

2.4.1 Summary and Personal Contribution

Zhou et al. 2023 provides novel molecular biology tools demonstrated to function in plant roots that allow researchers to exercise greater precision when engineering genetics for the purpose of study and application. The most commonly used promoters in plant genetic assays, such as 35S CaMV and nopaline synthase (NOS), cover less dynamic range and can be vulnerable to silencing when repeatedly used, possibly masking meaningful observations or limiting functionality of synthetic genetic circuits. My primary contributions were in protocol development and execution of validating conserved promoter function in dicots and root tissues utilizing *Medicago truncatula* via generating transgenic material through hairy root transformation.

2.4.2 Introduction

Modulation of transgene expression is crucial for plant bioengineering applications such as constructing gene circuits, plant metabolic engineering, or increasing efficacy of gene

editing efforts where the amount of protein being expressed by a synthetic promoter can drastically change outcomes. (388–391) Although a popular approach to tuning gene expression is the use of synthetic transcriptional systems, the lack of promoter choice limits engineering efforts to a single regime of universally high transgene expression strength and is an overall ineffective toolkit for modulating gene expression with a large dynamic range. The development of constitutive promoter libraries that drive a high dynamic range of expression strengths has been invaluable for synthetic biology efforts in most model microbial species. For example, the early development of such promoters in *Escherichia coli* through the systematic introduction of mutations on endogenous cis-elements has been widely utilized and improved the precision of experimental designs and metabolic engineering efforts in *E. coli.* (392,393) Thus, extending these approaches and the development of analogous plant parts would highly benefit the plant science community. Specifically, the model plant system *Nicotiana benthamiana* has emerged as an ideal platform for either rapidly prototyping designs or for biomanufacturing using advanced plant metabolic engineering. (394–396) Given the reliance on *N. benthamiana* by many plant synthetic biologists, the development of a suite of promoters with well-characterized gene expression strengths may provide the underlying tools to support a wide range of endeavors including pathway optimization, gene circuit designs, and genome engineering efforts.

There is significant interest in developing sophisticated genetic circuits in plants. Recent approaches for building gene circuits using either bacterial operators, trans-elements, or integrases fundamentally rely on plant promoter elements and precise expression strengths of these parts to ensure gene circuit function and to impart specificity to gene expression within the circuit. (391,397) One popular approach to tune gene expression is the use of orthogonal trans-elements, such as yeast Gal4 or a transcription-activation-like-effector (TALE), to target cis-elements to activate expression of a downstream transgene and has been demonstrated as an effective tool for transgenic protein expression in plant systems. (390,398,399) The complexities that afford tunability in such synthetic transcriptional systems may also lead to logistical challenges in construct design and implementation due to the requirement of introducing at least two components (e.g., the synthetic trans-element and synthetic promoter) to drive the expression of potentially only a single transgene. Thus, for simple applications, it may instead be more practical to use a well-characterized promoter with a defined expression level. Furthermore, a fully quantified library of plant promoters can still be utilized as a modular component for complex gene circuit designs such as tuning expression of a synthetic trans-element or supplementing baseline expression of an inducible promoter. (400,401) Finally, using the same promoter repeatedly to drive different transgenes can often lead to transgene silencing; thus, there is a need for alternative promoters with similar expression levels that can be used in plant engineering efforts. (402–404) Such issues are made more prominent given the dearth of choice for plant constitutive promoters resulting in promoter designs primarily using CaMV 35S enhancer, nopaline synthase promoter variants, or ubiquitin and actin cis-regulatory elements. (405–409) Altogether, the development of a library of self-contained constitutive promoters that can express transgenes at various well-defined

strengths would simplify construct design, enable faster pilot experiments in plant systems, and provide a versatile tool for plant synthetic biologists.

In this study, we developed a promoter library with well-defined gene expression strengths and consistent activity in multiple plant systems. Our library encompasses 15 promoters, which span two orders of magnitude of expression strength with 5 promoters on par with 35S promoter activity in *N. benthamiana* transient expression assays. We extensively characterize the expression of transgenes driven by these promoters in terms of biological variation and transgene identity. We evaluate the efficacy of using these promoters as internal standards for normalization of transgene expression in experimental contexts. Finally, we demonstrate that PCONS promoter activity is conserved in diverse plant species, such as *Lactuca sativa* using self-transcribing active regulatory region sequencing (STARR-seq) and can drive transgene expression in *Medicago truncatula* via hairy-root transformation. Overall, these findings enable the adoption of the PCONS promoter library for use in any application where precise control of transgene expression strength is a necessity.

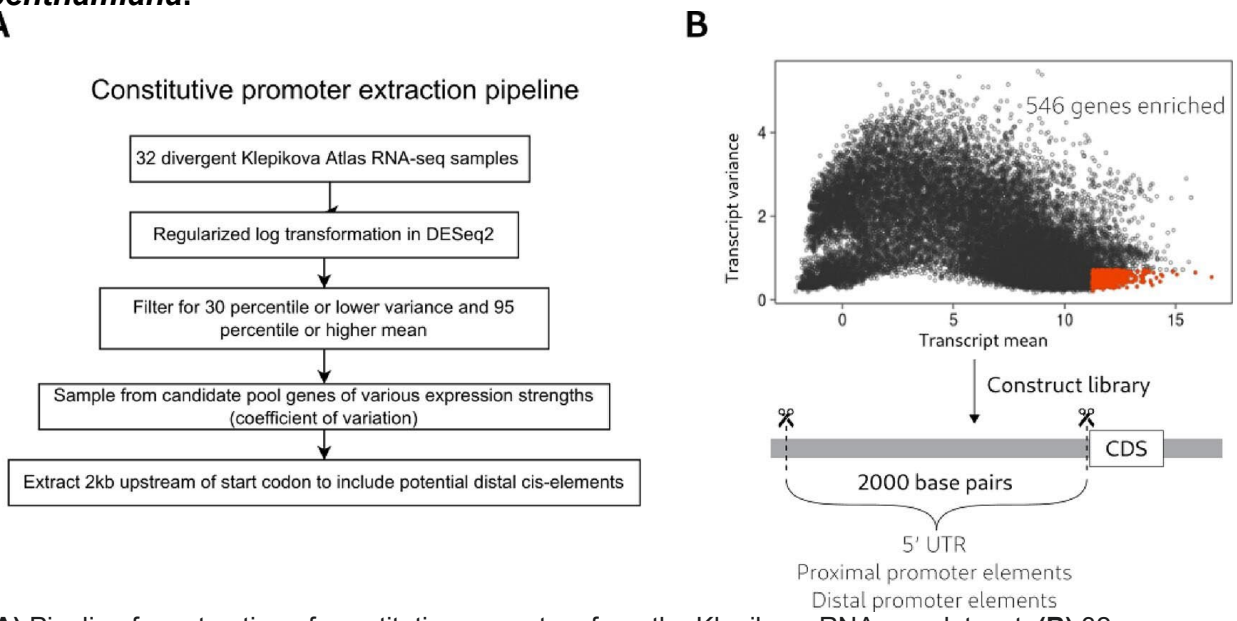
2.4.3 Results

2.4.3.1 Mining and Validation of a Constitutive Promoter Library from Publicly Available RNA-Seq Datasets

We envisioned the development of a plant promoter library with (1) high dynamic range, (2) ubiquitous expression in different tissues, and (3) activity in different plant species. With the abundance of RNA-seq datasets measuring Arabidopsis mRNA abundance spanning a wide array of tissues, organs, and developmental stages, we sought to integrate this information into a computational pipeline for extracting promoters displaying consistent expression activity across all tissues. In our promoter discovery pipeline, we emphasized filtering candidate promoters based on mean and variance gene expression quantiles across samples to identify and isolate promoters for a given plant transcriptomics dataset with consistent gene expression with respect to tissues and/or plant development (Figure 2-8A). We applied this pipeline to a widely-used Arabidopsis developmental transcriptomics atlas by Klepikova et al., which evaluated 79 organs and developmental stages of Arabidopsis, in order to demonstrate the efficacy of the pipeline to identify ubiquitously expressed promoters in different plant organs and developmental space and to build and validate a ready-to-use promoter library for many dicot systems. (410) We selected a subset of tissue samples from the Klepikova atlas, which were largely uncorrelated in terms of expression of each individual gene (Supplementary Figure 2-9) in order to maximize our candidate pool coverage of gene expression variance with respect to plant development and tissue type. By maximizing the gene expression variance space and subsequently filtering and constraining this space for low gene-expression variance promoters within the dataset, we enrich promoter designs, which likely display stable expression patterns with respect to plant development and tissue type (Figure 2-8B). Additionally, these enriched promoters may contain an abundance of cis-regulatory elements tuned for constitutive gene expression, which may be conserved and thus are more likely to display desirable activity in other plants. mRNA abundances obey Zipf's law, where a few genes account for the majority of gene expression.(411) Given this relationship between gene rank and mRNA

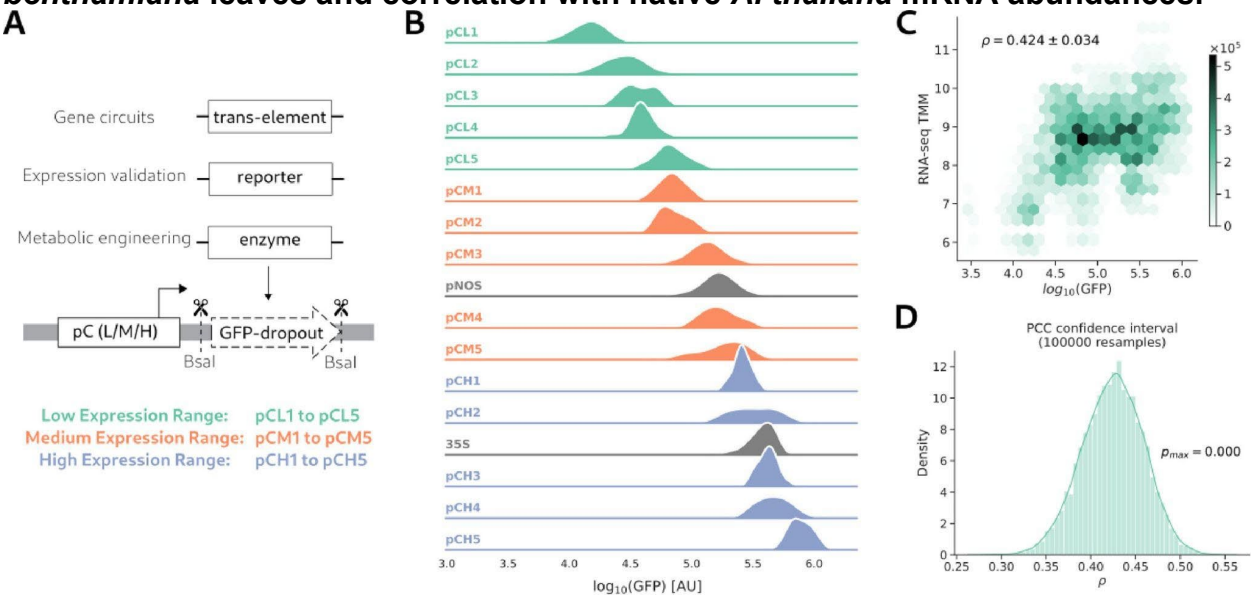
abundance, we selected for the top 5 percent of genes with respect to mean expression, which encompassed approximately 44% of all mapped reads in our compiled dataset thus enriching for a range of potential promoters with a wide but measurable dynamic range. In total, genes were subject to quantile filters selecting for high (>95 percentile) mean expression and low variance in expression (<30 percentile) resulting in 546 candidate genes (~2% of total genes) in Arabidopsis with sufficiently strong and spatially constitutive expression activity detectable in downstream characterization assays (Figure 2-8B).

Figure 2-8. Data pipeline for identifying constitutive promoters from RNA-seq datasets and validation of promoter activity through transient expression in *N. benthamiana*.



(A) Pipeline for extraction of constitutive promoters from the Klepikova RNA-seq dataset. **(B)** 32 regularized RNA-seq samples mapped by mRNA abundance mean and variance. Red highlighted regions of interest are quantile filtered genes by mean (>95%) and variance (<30%).

Figure 2-9. GFP expression output of PCONS promoters in agroinfiltrated *N. benthamiana* leaves and correlation with native *A. thaliana* mRNA abundances.



(A) Design of GFP-dropout plasmid for convenient swapping of transgenes driven by a PCONS promoter. **(B)** Ridge plot displaying mEGFP expression of the PCONS library after agroinfiltration of constructs into *N. benthamiana* leaves (n = 3). **(C)** Correlation between normalized mRNA abundance in TMM (trimmed mean of m values) of PCONS genes and GFP expression driven by PCONS promoters. **(D)** Confidence interval of Pearson correlation coefficient (PCC) between GFP fluorescence and RNA-seq gene counts (as TMM) from the Klepikova RNA-seq dataset as extrapolated via bootstrapping.

To build a suite of constitutive promoters with a wide range of transgene expression strengths as a plant synthetic biology toolkit, we selected 15 genes spanning different mean abundance of transcripts from the enriched candidate pool (Table 2-1). We extracted a region of interest starting two kilobases upstream of the start codon to the beginning of the coding sequence of each gene to ensure capture of cis-regulatory elements responsible for their respective transcriptional activity. (412–414) Promoters were inserted into a pCAMBIA backbone upstream of PhytoBrick compatible GFP dropout cassette for facile construction of plasmid designs for metabolic engineering or plant synthetic biology (Figure 2-9A). This simple cloning strategy allows for the GFP dropout cassette to be replaced by the transgene of interest; thus, bacterial colonies with successful construction of the plasmid that have lost the GFP dropout cassette no longer fluoresce green and can be easily screened for on an agar plate. We sampled the transgene expression of these 15 promoters driving a reporter GFP and measured fluorescence intensity in *N. benthamiana* leaf disks after agroinfiltration. The full suite of promoters exhibited expression levels spanning nearly two orders of magnitude (Figure 2-9A). Notably, seven of our promoters achieve expression strengths that are stronger than the NOS promoter, and five are on par with the 35S promoter. From these GFP measurements, we divided the 15 constitutive promoters into three groups denoted as low (PCL), medium (PCM), and high (PCH) and denoted the ensemble of promoters as the PCONS library. PCONS promoter strength measured through reporter gene expression was positively correlated (ρ = 0.424) with Klepikova atlas RNA-seq readouts (Figure 2-9B, C).

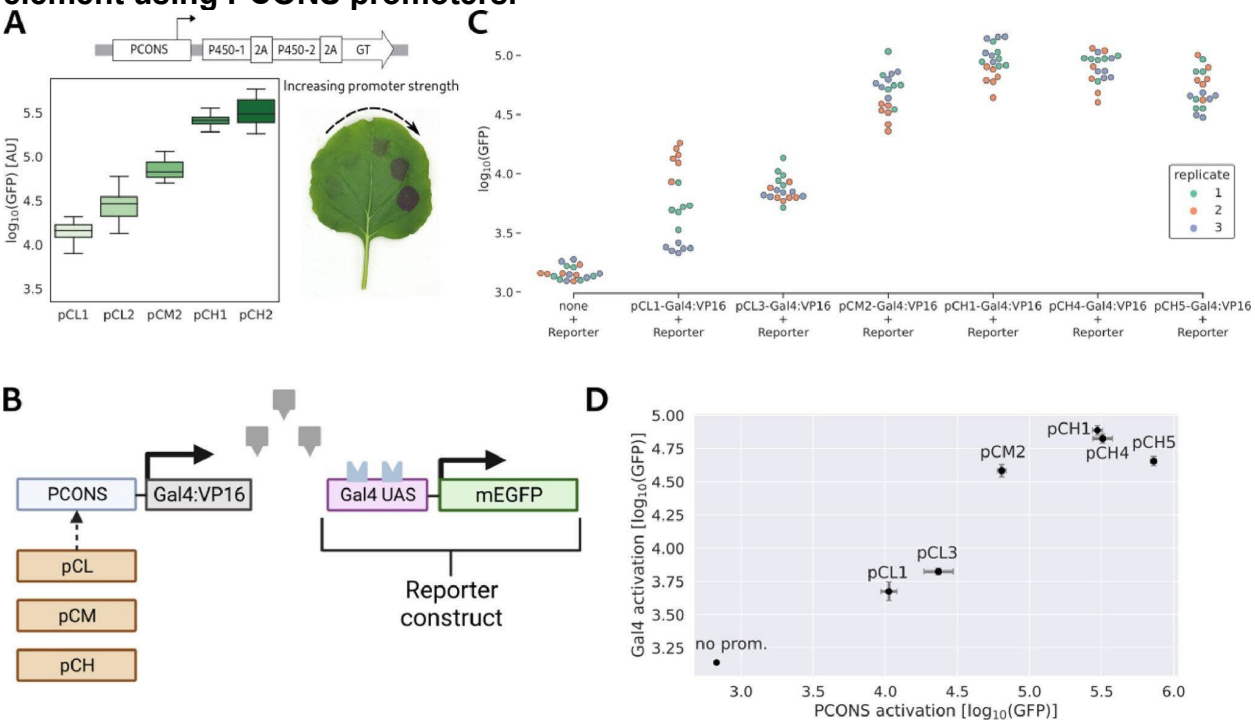
Table 2-1. PCONS Library TAIR IDs and Putative Gene Functions.

PCONS ID	TAIR ID	gene identity
pCH1	AT1G23490	ADP-Ribosylation factor 1A
pCH2	AT1G07890	ascorbate peroxidase 1
pCH3	AT4G13930	serine hydroxymethyltransferase 4
pCH4	AT1G15690	AT V-PPase 3
pCH5	AT5G11740	Arabinogalactan protein 15
pCM1	AT2G44060	late embryogenesis abundant 26
pCM2	AT4G40040	HTR5 histone
pCM3	AT3G15730	phospholipase D alpha 1
pCM4	AT3G13920	eukaryotic translation initiation factor 4A-1
pCM5	AT3G55440	cytosolic isoform triose phosphate isomerase
pCL1	AT1G76200	NADH dehydrogenase ubiquinone 1 beta subcomplex unit
pCL2	AT2G33040	gamma subunit of Mt ATP synthase
pCL3	AT4G34110	arabidopsis poly(a) binding 2
pCL4	AT1G22300	general regulatory factor 10
pCL5	AT3G44110	dnaj homologue 3

2.4.3.2 Titratable Activation of Transgene Expression Using PCONS

Agroinfiltration in *N. benthamiana* is a gold standard methodology for plant metabolic engineering. One of the most popular plasmid systems used for *N. benthamiana*-based metabolic engineering efforts is the pEAQ plasmid system, which uses 35S and cowpea mosaic virus UTRs to enable strong levels of transgene expression. (409) To further validate the dynamic range of the PCONS promoter library, we demonstrate distinct and incrementally increasing gene expression of the RUBY reporter system with 5 different PCONS promoters into the same *N. benthamiana* leaf. (415) We observed increasing red coloration of the infiltrated spots as PCONS promoter strength increased (Figure 2-10A). The gradation of betalain production observed in *N. benthamiana* leaf presents a direct demonstration of modulating metabolic flux, displays PCONS as distinct modules with consistent expression strengths for synthetic biology, and supports the usage of PCONS promoters as a potential tool for metabolic engineering.

Figure 2-10. Stepwise expression of a metabolic pathway and trans-activator element using PCONS promoters.



(A) Selection of 5 promoters used to express the betalain biosynthesis pathway with increasing strength in a single *N. benthamiana* leaf. Boxplots show GFP expression strengths of promoters (n = 3) used for betalain biosynthesis. Betalain is visualized but production was not quantified. **(B)** Visualization of mGFP expression driven by a Gal4 UAS using PCONS for regulation of Gal4:VP16 and **(C)** measurements using different PCONS upstream of Gal4:VP16. **(D)** Comparison of average fluorescence readout driven directly through PCONS promoters versus indirectly driven through expression of Gal4:VP16. Error bars represent standard error of mean.

Beyond metabolic engineering, the development of synthetic transcriptional systems and genetic circuits in plants has been of great interest. Orthogonal trans-elements, such as the yeast Gal4 transcription factor, are often utilized as a binary switch to toggle expression of a target gene and are the lynchpin component in recent efforts to build complex circuits in plants. (391,398) As such, the ability of trans-regulatory elements in activation of expression of a target gene is often viewed as strictly on/off depending on the presence or absence of the element in a given system. However, gene expression could also be conceptualized as a function where the rate of mRNA production depends on the concentration of TF input and thus further granularity of gene expression activation can be achieved by changing the expression level of a given trans-element through an upstream promoter. (416) Trans-elements have a concentration-dependent dynamic range in their ability to activate gene expression, and within this range, tunable expression can be achieved exclusively through modulating trans-element concentration. To uncover the dynamic range of activation for a popular activating trans-element, Gal4:VP16, we used six promoters from our PCONS library to span a range of expression strengths to modulate its expression (Figure 2-10B). By measuring a downstream GFP reporter with Gal4 upstream activation sequences (UAS) as binding sites, we observe incremental activation of GFP with increasing expression of

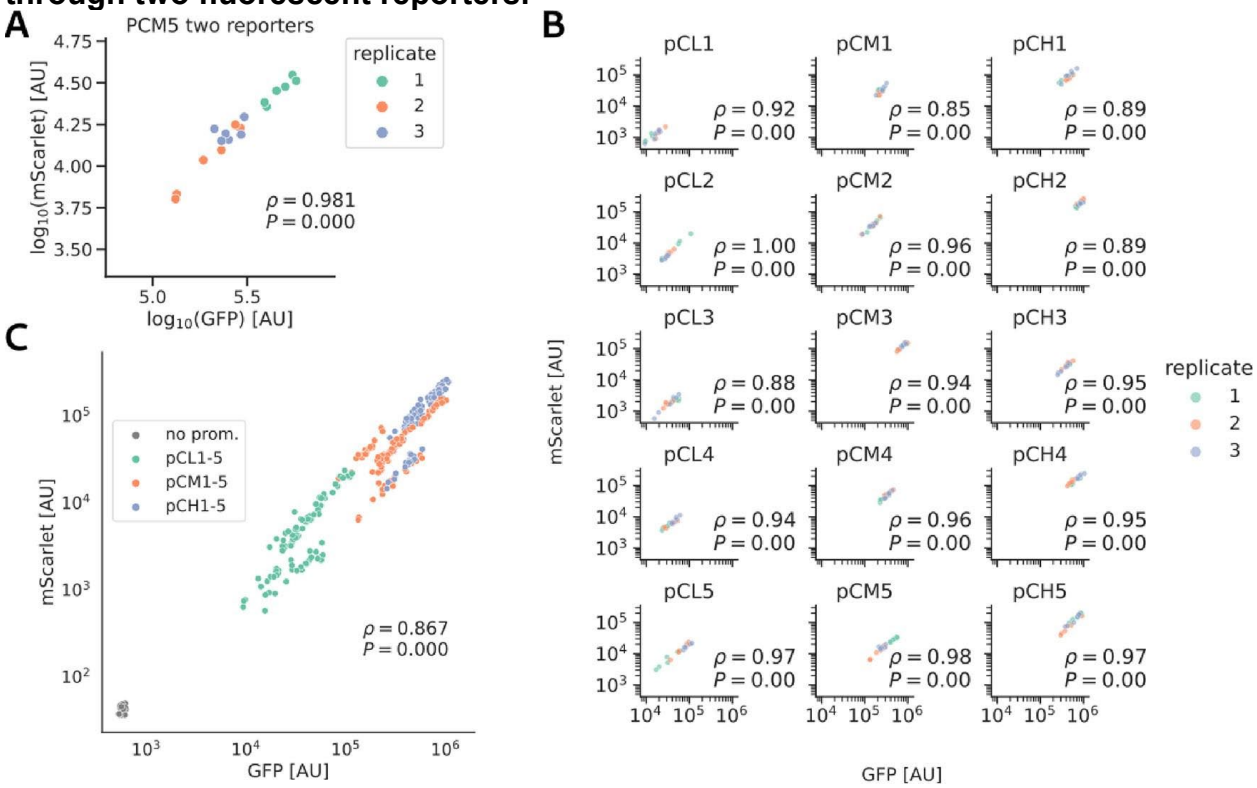
Gal4:VP16 (Figure 2-10C). Notably, the reporter signal is saturated when using medium to high promoter strengths (pCM2-pCH5). The dynamic range of GFP fluorescence is observed only when using weaker promoters (pCL). When these GFP readouts are framed with respect to previous PCONS measurements (Figure 2-9A) as a proxy for Gal4:VP16 concentration (Figure 2-10D), the data recapitulates a binding curve where the GFP readout saturates when nuclear Gal4:VP16 concentration is much higher than the Gal4 KD. Thus, tuning the activation strength of Gal4:VP16 is achieved exclusively using weaker promoters within the PCONS library and further emphasizes the need for synthetic parts with diverse expression strengths necessary to access tunability and dynamics for a given transgene expression platform.

2.4.3.3 PCONS Show Consistent and Correlated Transgene Expression of Two Fluorescent Reporters

Shallow and deep machine learning models have demonstrated that gene coding sequences (CDS) and mRNA stabilities have strong predictive power on the overall expression of a given gene, thus suggesting that the CDS context could potentially dictate promoter activity. (417) In order for the PCONS promoter library to be an effective toolkit for plant synthetic biology, strong PCONS promoters should remain strong relative to weaker PCONS promoters regardless of CDS content. Thus, we aimed to establish whether the relative activity between different PCONS promoters can change due to transgene identity.

We tested this facet of PCONS through co-infiltration of two reporter fluorescent proteins (GFP and mScarlet) driven by the same promoter for each respective PCONS promoter. Visualizing all PCONS promoters individually driving two reporter proteins in the same plot, we observed positive correlation between expression of GFP and mScarlet (Figure 2-11A,B). Each individual promoter displayed correlated expression strengths between mScarlet and GFP for both technical and biological replicates and formed clusters according to biological replicates (Figure 2-11A). These trends suggest that within a given single promoter, transgene expression strength is consistently maintained regardless of differences in transgene identity, plant state, and technical error for the entire PCONS library (Figure 2-11B). While the rank order of promoter strengths is not exactly the same when using PCONS to drive GFP versus mScarlet, promoter activity is positively correlated in expressing each transgene ($\rho = 0.862$), although certain promoters may favor expression of one fluorescent protein over another (Figure 2-11C). This correlated gene expression activity corroborates the robustness of PCONS promoters in driving expression of two fluorescent reporters.

Figure 2-11: PCONS-driven transgene expression bias and variation tracked through two fluorescent reporters.



(A) Expression of mEGFP and mScarlet was simultaneously measured after co-infiltration of two constructs containing PCM5 promoter driving expression of each fluorophore. **(B)** Co-infiltration of two fluorescent reporters driven by each PCONS promoter. **(C)** Correlation of PCONS promoter activity when simultaneously driving two different fluorescent transgenes.

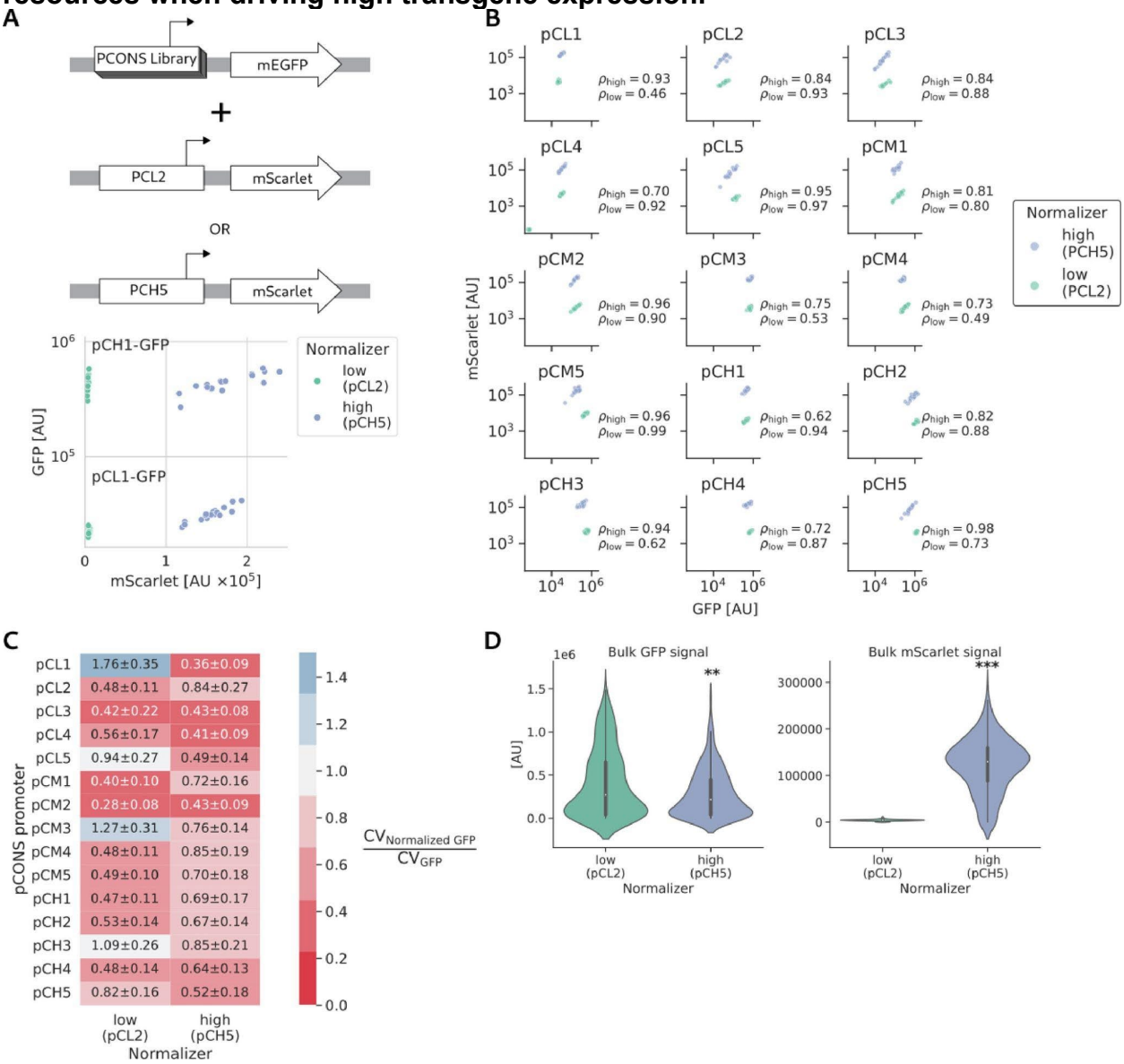
2.4.3.4 Optimizing PCONS for Signal Normalization in Plant Gene Expression Studies

Normalization of signals is fundamental in the design of reproducible biological experiments by increasing precision of measurements and reducing certain facets of systematic and biological error affecting both the gene of interest and the normalizing gene. In the context of quantifying plant gene expression using agroinfiltration, many different promoters, reporters, and construct designs utilize dual luciferase or dual fluorescence reporters for signal normalization. Despite the popularity in the use of a second reporter to enable greater precision in readouts for promoter strength, empirically identifying ideal constructs for normalization has been less well-studied. Through evaluation of PCONS promoters as an internal standard we observed: (1) the reduction in the noisiness of signal quantified by the coefficient of variation and (2) presence of a maximum capacity for transgene expression caused by using overtly strong PCONS promoters for signal normalization.

We evaluated the robustness of normalization by infiltrating individual PCONS promoters driving expression of GFP and co-infiltrating a choice between weak or strong PCONS promoters (PCL2 and PCH5, respectively) to drive expression of mScarlet to normalize GFP fluorescence readouts. Replicate variation is spread out into

different dynamic ranges of the mScarlet signal depending on the choice of PCONS driving mScarlet expression (Figure 2-12A). The fluorescent readout was correlated between the mScarlet and GFP channels when using either low or high PCONS strength for normalization and with positive covariance overall between the two channels (Figure 2-12B), suggesting that each fluorescent channel captures a correlated biological variation, ultimately corrected through normalization, and thus boosting the resolution of differences between each individual PCONS promoter. To quantify the efficacy of normalization on reducing the experimental error, we determined the coefficient of variation (CV) or relative standard deviation of each PCONS promoter for both the raw GFP signal and normalized signal. After normalization, the coefficient of variation for each promoter generally decreases (CV ratio < 1) demonstrating the efficacy of using PCONS promoters to normalize gene expression measurements (Figure 2-12C). In specific cases, the coefficient of variation increases after normalization, suggesting additional interaction between the promoter used for biological signal measurement and the promoter used for normalization. Therefore, the effect of normalizing on the coefficient of variation of the measured signal must be considered when designing experiments with internal standards. Overall, the reduced CV in normalized signal suggests that PCONS promoters can be an effective choice for an internal standard for gene expression studies.

Figure 2-12: Efficacy of normalization using PCONS and limitation of cellular resources when driving high transgene expression.



	low (pCL2)	high (pCH5)
pCL1	1.76±0.35	0.36±0.09
pCL2	0.48±0.11	0.84±0.27
pCL3	0.42±0.22	0.43±0.08
pCL4	0.56±0.17	0.41±0.09
pCL5	0.94±0.27	0.49±0.14
pCM1	0.40±0.10	0.72±0.16
pCM2	0.28±0.08	0.43±0.09
pCM3	1.27±0.31	0.76±0.14
pCM4	0.48±0.11	0.85±0.19
pCM5	0.49±0.10	0.70±0.18
pCH1	0.47±0.11	0.69±0.17
pCH2	0.53±0.14	0.67±0.14
pCH3	1.09±0.26	0.85±0.21
pCH4	0.48±0.14	0.64±0.13
pCH5	0.82±0.16	0.52±0.18

(A) GFP expression of PCL1 and PCH1 promoters using two different PCONS promoters (PCL2 or PCH5) to normalize the GFP signal through mScarlet expression. **(B)** Correlation of two fluorescent channels when using two (PCL2 or PCH5) different promoters driving mScarlet to normalize the entire PCONS library driving GFP. **(C)** Bootstrap estimation of the ratio of coefficient of variation for normalized PCONS-driven GFP expression vs unnormalized PCONS-driven GFP expression. **(D)** Bulk GFP signal for the PCONS library driving GFP when using two different promoters (PCL2 or PCH5) for normalization. Significant changes determined through the Kruskal–Wallis test H-test.

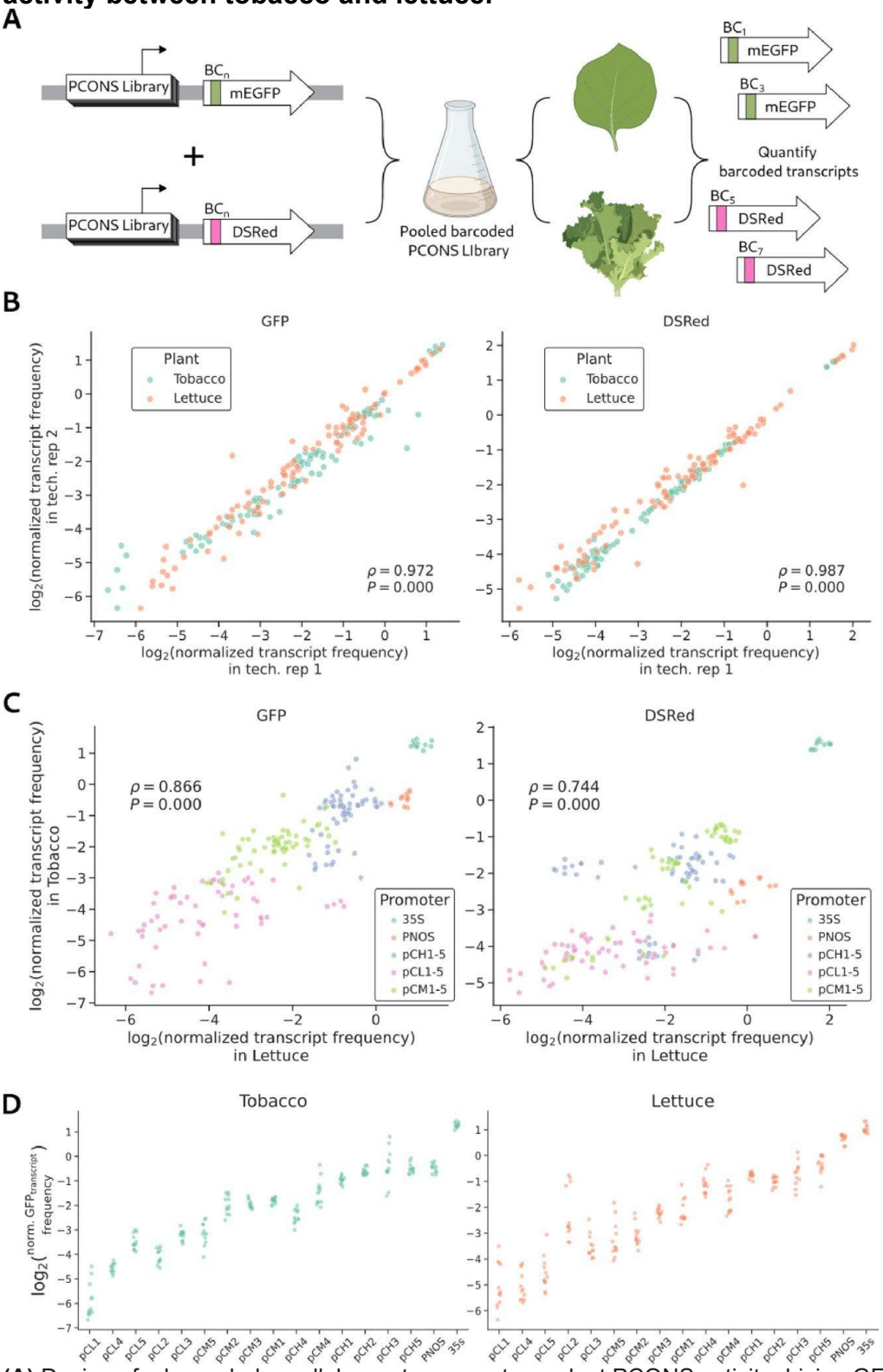
In a wide range of microbial studies, resource allocation has been demonstrated to affect the physiology and overall metabolism of biological systems. (418–420) The introduction of strong exogenous promoter elements driving expression of transgenes imparts a metabolic burden and limits cellular resources. While normalization with low or high PCONS promoters provides a reduction in measurement noise, we observed significant attenuation of the bulk GFP signal when using the strongest PCONS

promoter (PCH5) for normalization versus a weaker promoter (PCL2) (Figure 2-12D). This finding suggests a maximum capacity for transgene expression in a plant system, which can be a limiting factor depending on user experimental goals. Our results shed light on how transient expression studies in *N. benthamiana* that simultaneously express multiple genes at high expression levels may impact transcriptional levels and potentially the overall metabolic state of transformed plant cells. These effects should be accounted for through the careful selection of promoters, highlighting the utility of the PCONS library.

2.4.3.5 PCONS Library Displays Conserved Activity in Different Dicot Systems

To validate the general activity of the PCONS library in different plant systems, we quantified transcriptional activation of each promoter using a STARR-seq assay (Figure 2-13A). (421,422) We generated a mixed plasmid library of PCONS promoters driving GFP or DsRed CDSs containing an 18 nt barcode introduced via PCR. We subsequently agro-infiltrated the library into *N. benthamiana* and *L. sativa* and through next-generation sequencing of the resultant barcoded RNA transcripts obtained simultaneous readout of promoter activity in driving expression of each transgene. Fold-enrichment of both GFP and DsRed transcripts driven by PCONS promoters share similar dynamic range in both *N. benthamiana* and *L. sativa* (Figure 2-13B). Furthermore, we observed a positive correlation in transcriptional activation of GFP and DsRed between both plant systems, suggesting that PCONS promoter activity is conserved despite millions of years of evolutionary divergence between each clade (Figure 2-13C). The GFP transcript abundance for each PCONS promoter is generally consistent with protein level measurements based on fluorescence measurements (Figure 2-13D). The 35S promoter displays the strongest transcriptional activation activity despite lower fluorescence readouts when compared with high activity PCONS promoters (Figure 2-9A). The high transcription activation strength of 35S could potentially incur a detrimental consequence on production of fluorescent protein aligned with our observation of an attenuation of fluorescence readout when using strong promoters for both readout and normalization (Figure 2-13D), whereas usage of a single strong PCONS promoter may not yet incur this penalty.

Figure 2-13: STARR-seq assay reveals conservation of PCONS transcriptional activity between tobacco and lettuce.

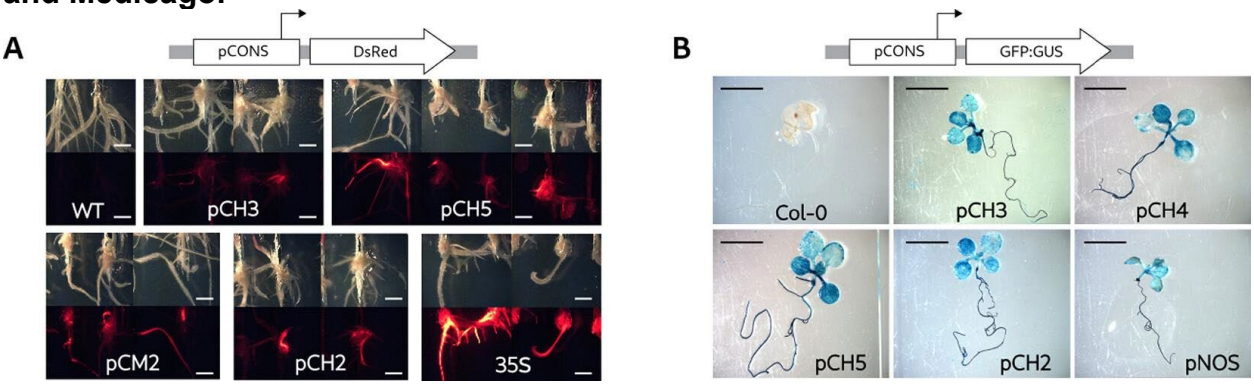


(A) Design of a barcoded parallel reporter assay to readout PCONS activity driving GFP and DsRed transgene expression in *N. benthamiana* and *L. sativa*. **(B)** Pearson correlation of technical replicates and

shared dynamic range for PCONS driving dual transgenes in *N. benthamiana* and *L. sativa*. Each data point represents the median count of a given PCONS promoter inside a biological replicate and all PCONS promoters are present in the data. **(C)** Pearson correlation of transgene expression activation by PCONS promoters between *N. benthamiana* and *L. sativa*. Promoters are binned to low (PCL1 to PCL5), medium (PCM1 to PCM5), and high (PCH1 to PCH5). Each data point represents the median count of a given PCONS promoter inside a biological replicate, and all PCONS promoters are present in the data. **(D)** Activation strength of individual PCONS in transcribing GFP. Each data point represents the median count of a given PCONS promoter inside a biological replicate and all PCONS promoters are present in the data.

Additionally, we wanted to confirm that the conserved PCONS activities observed in transient assays are maintained in stable integrations. Thus, we generated stable integrations in A. thaliana via Agrobacterium floral dipping for 5 different promoters in the PCONS library driving a GFP-GUS fusion protein. As expected, beta-glucuronidase (GUS) activity was visualized throughout transgenic Arabidopsis seedling root, stem, and leaves (Figure 2-14B). Conserved promoter activity across different plant systems is a common feature of popular promoters used for plant synthetic biology. To explore the efficacy of the PCONS library in driving expression in another dicot, we performed hairy root transformation of *Medicago truncatula* to express DsRed driven by PCONS promoters (Figure 2-14A). Transformed roots expressed DsRed with levels comparable to 35S promoter though, visually, 35S expressed DsRed strongest in Medicago. These results suggest that PCONS promoter activities, mined from Arabidopsis transcriptomics, can translate to other dicot systems such as *M. truncatula*, *N. benthamiana*, and *L. sativa*. Thus, the selected PCONS promoters demonstrate potential to increase the promoter options for stable plant transformation and can circumvent potential gene silencing incurred by using the same promoter to express multiple transgenes.

Figure 2-14: PCONS promoters drive stable transgene expression in Arabidopsis and Medicago.



(A) Hairy root transformation of Medicago expressing DsRed driven with various PCONS promoters. White bars indicate 5 mm. **(B)** GUS staining of T2 Arabidopsis seedlings with stable integrated GFP:GUS expression cassettes driven with various PCONS promoters. Black bars indicate 1 cm.

2.4.4 Discussion

New high-throughput approaches in plants have resulted in an unprecedented amount of potential parts for plant synthetic biology. Validation of gene expression at the protein level, benchmarking newly discovered parts in relation to standardized parts, and

expanding the dynamic range of biological activity are crucial for enabling successful experimental designs and engineering efforts in planta. We present a library of plant promoters quantified at both RNA and protein level in the expression of a variety of transgenes compared directly to popular gold-standard plant constitutive promoters. We demonstrate the utility of the PCONS library in incrementally increasing expression of biosynthetic enzymes for betalain synthesis and tunable gene activation using Gal4:VP16. We dimensionalize technical and biological variation and quantify the efficacy of normalization using different PCONS promoters, thus demonstrating their potential as internal standards for normalizing biological variation. PCONS promoters displayed correlated transcriptional activity in both *L. sativa* and *N. benthamiana* and stable expression of GUS and DsRed in A.thaliana and M. truncatula, respectively, demonstrating their widespread activity in multiple plant systems. Overall, PCONS promoter activity was enriched *in silico* for constitutive gene expression through tissues in Arabidopsis development using appropriate transcriptomics data, and these findings bolster the benefits of transcriptomic screening for engineering features in plant promoters. (400) However, a full empirical demonstration of constitutive activity of the PCONS library over plant development has yet to be demonstrated.

The dynamic range of transgene expression enforces potential limits to the tunability for a given parts library. An exhaustive transcriptional assessment of all Arabidopsis core promoter elements using a massively parallel reporter assay (MPRA) spanned approximately a 2 (9)-fold change at the transcript level. (423) In this study, the PCONS library displays a 2 (6)-fold enrichment of transcripts and protein between weakest and strongest PCONS promoters. Given that the PCONS promoters encompass the top 5% of expressed genes on average in A. thaliana, unenriched endogenous promoter sequences are likely to be weaker in expression strength. Thus, the dynamic range of PCONS could expand with the inclusion of promoters extracted from more weakly expressed genes; however, gene expression of these promoters may be harder to quantify if sufficiently weak in strength, and their utility in driving expression of transgenes may be limited. Ultimately, our data pipeline utilizing the Klepikova Arabidopsis development RNA-seq dataset was able to identify promoters, which demonstrate consistent and generalizable performance in two unique dicot systems and the highly detectable dynamic range in PCONS promoters could provide utility as candidate reference genes for quantitative PCR in Arabidopsis. We envision that our mining approach could apply analogously to other transcriptomics-rich plant systems such as rice and maize to further expand on the available compatible synthetic promoter scaffolds for engineering monocots.

Machine learning models for gene expression in yeast suggest that different cis-regulatory components can each have unique contributory effects on the overall gene output. (417) Indeed, successful strategies in tuning gene expression for plant synthetic biology take advantage of these regulatory facets by modularizing these cis-elements to achieve enhanced dynamic range and fine-tuned levels of transgene expression. (399,424) High levels of gene expression, comparable with 35S, can be achieved by leveraging orthogonal trans-element activators and patterning plant cis-regulatory elements into synthetic promoters. For the design strategy of the PCONS library, we

took a sequence agnostic approach to extract 2 kb upstream of the start codon of each gene regulated by a respective PCONS promoter. Therefore, the granular cis-regulatory information, which contributes to the overall promoter activity of PCONS, some of which lie in intergenic regions, remains experimentally unexplored. Thus, we envision the PCONS library as a useful tool that can be utilized in whole or piecemeal combinations with previously developed cis-regulatory parts and strategies to further enrich the existing suite of tools for plant synthetic biology.

Plants are capable of natural synthesis of a variety of complex compounds, and modulation of metabolic flux for these endeavors requires fine-tuned expression of multiple transgenes to maximize production of useful biological products. A common preliminary strategy in metabolic engineering efforts entail overexpression of multiple biosynthetic enzymes using strong promoters, which may result in overaccumulation of intermediate compounds. (425–427) Our findings explore the downstream consequences of using strong promoters to drive multiple transgenes. Using two fluorescent reporters (GFP and mScarlet), we found our strongest PCONS promoter (PCH5) attenuated GFP fluorescence when tasked with simultaneous production of both transgenes. We speculate that this effect demonstrates a cellular limit in protein expression capacity when using strong promoters to drive multiple transgenes. To address these potential challenges, future metabolic engineering efforts may be able to utilize the PCONS library to tune metabolic flux, decrease overall transcriptomic load on plant cells, and circumvent potential engineering pitfalls. Overall, the PCONS library offers plant scientists a simple, streamlined plasmid expression system with a selection of well-defined promoter strengths for transgene expression.

2.4.5 Methods

2.4.5.1 Data Pipeline for Identifying and Analyzing Promoters for the PCONS Library

32 highly divergent samples from the Klepikova atlas were identified using sample correlation values from (410) (Supplementary Table 2-1). Klepikova RNA-seq BAM files were obtained from CyVerse Data Commons (https://datacommons.cyverse.org/) and converted into count data using featureCounts. (428) Read data is normalized using the regularized log (rlog) and variance stabilizing (VST) transformations (DESeq2 (429)) and quantile filtered for >95%-tile mean gene expression and <30% variance of gene expression. Enriched genes from both transformations are combined and arranged by their coefficient of variation (variance/mean), resulting in 546 unique genes. 15 genes of differing coefficient of variation were chosen for downstream evaluation and construction of the PCONS library. Bootstrapping of the Pearson correlation coefficient was performed between Klepikova atlas trimmed mean of m values (edgeR (430)) and measured GFP fluorescence of PCONS promoters from this study. Specifically, TMM values were calculated and aggregated for 32 samples in the Klepikova atlas for each gene regulated by a PCONS promoter. 18 random pairs of data, RNA-seq TMM and fluorescence, were sampled with replacement for each PCONS promoter for 10,000 iterations to a confidence interval for the Pearson correlation coefficient between the two datasets.

2.4.5.2 Plasmid Construction

All constructs are based off of the pCAMBIA plasmid backbone. PCONS promoter sequences were PCR amplified from A. thaliana gDNA with and assembled into a pCAMBIA2301 backbone through Gibson assembly using NEBuilder HiFi Assembly Mix (NEB) and kit protocols (Tables 2-2 and 2-3). A GFP dropout cassette with common syntax BsaI sites was inserted downstream of each promoter resulting in pC(L/M/H)(1-5)-GFPdropout plasmid for convenient subsequent cloning steps. (431) Gal4:VP16, mScarlet, DsRed, and mEGFP were introduced into each PCONS-GFP dropout construct using Type IIS restriction cloning described in the previous literature. (432) All constructs and plasmid maps are publicly available on the JBEI Inventory of Composable Elements (https://public-registry.jbei.org/) (Table S4).

2.4.5.3 Leaf Punch Assay

PCONS plasmids were transformed into GV3101 *Agrobacterium tumefaciens* via electroporation. Agroinfiltration protocol was adapted from Sparkes et al. (433) Transformed agrobacterium were grown in LB liquid media with 50 µg/mL kanamycin, 50 µg/mL rifampicin, and 30 µg/mL gentamicin to between optical density (OD) 0.6 and 1 before diluting to OD 0.5 in agroinfiltration buffer (10 mm MgCl2, 10 mm MES, pH 5.6). *N. benthamiana* plants were grown and maintained in a temperature-controlled growth room at 25 °C and 60% humidity in 16/8 h light/dark cycles with a daytime PPFD of ~120 µmol/m2 s. *N. benthamiana* were germinated and grown in Sungro Sunshine Mix #4 supplemented with ICL Osmocote 14-14-14 fertilizer at 5 mL/L and agroinfiltrated at 29 days of age. Constructs of interest were infiltrated into the fourth leaf (counting down from the top of the tobacco plant) and harvested 3 days post-infiltration. Eight leaf disks per plant were extracted and arranged atop 250 µL of water in black, clear-bottom, 96-well microtiter plates (Corning). Measurements of GFP and RFP fluorescence were recorded for each leaf disk using a BioTek Synergy H1 microplate reader (Agilent). Technical replicates were uniformly rank-filtered to the top 6 of 8 disk readouts for all conditions to eliminate empty leaf disk readouts. The code used for analysis of fluorescence data is available at https://github.com/shih-lab/PCONS_Analysis.

2.4.5.4 STARR-Seq Library Design

The PCONS STARR-seq library was generated using the pPSup (https://www.addgene.org/149416/) plasmid as template replacing the coding sequence, and terminator regions with GFP or DSRed, and tUBQ3 terminator, respectively. The resultant plasmid was barcoded through PCR followed by introduction of each PCONS using Golden Gate assembly and transformed into *E. coli* (XL1-Blue). Approximately 50 E.coli colonies of each pCONS construct harboring GFP or DsRed were mixed into a single pooled library. Plasmids were isolated for barcode association using PacBio Sequel II sequencing (Azenta). The complete PCONS library was subsequently transformed into *A. tumefaciens* (GV3101) harboring the pSoup helper plasmid through electroporation. (434)

2.4.5.5 STARR-Seq Assay

30 mL of Agrobacterium harboring the PCONS promoter library was grown to an OD of 1.0, and subsequently 5 mL was set aside as an input sample after plasmid isolation with

a Plasmid Plus Midi Kit (Qiagen). The PCONS promoter library was introduced into six independent tobacco and lettuce plants through leaf tissue using agroinfiltration. Tissues were subject to RNA extraction 3 days post-infiltration using an RNEasy Plant Mini Kit (Qiagen). Two technical replicates for cDNA synthesis were performed for each RNA sample using complementary primers specific for either GFP or DsRed using SuperScript IV Reverse Transcriptase (Thermo-Fisher). Barcoded forward primers (Table S2) were used for cDNA amplification for downstream demultiplexing of biological replicates after optimal amplification cycles were identified using qPCR (SYBR Green Master Mix, BioRad). Sequencing adapter ligation, TapeStation, and Bioanalyzer quality control of cDNA amplicons and 2× 150 paired-end sequencing was performed by Azenta Life Sciences using an Illumina HiSeq 4000.

2.4.5.6 Analysis of STARR-Seq

In order to identify PCONS promoters from barcoded reads, a dictionary between PCONS promoters and barcodes was constructed from PacBio Sequel II sequencing of the input PCONS promoter library. Subreads were processed using the PacBio ccs package into an unaligned BAM file and subsequently aligned with PCONS promoter sequences. The aligned read barcodes and PCONS identities were then processed into a dictionary object in Python for downstream analysis. Paired-end sequencing data from plant cDNA barcoded amplicons were in mixed orientation due to ligation of sequencing adapters. Paired-end reads were combined using PANDAseq. (435) Cutadapt was used for adapter trimming and to reorient and demultiplex reads into their respective biological replicate and transgene identities. (436) Barcodes with less than five reads were discarded for downstream analysis. Each barcode frequency (barcode count divided by all counts) was mapped to their respective promoter identities, and enrichment of each barcode was calculated by dividing by the barcode frequency in the input sample. The average enrichment of all promoters was then calculated and visualized for all samples. Scripts used for processing and analysis of sequencing data are available at https://github.com/shih-lab/PCONS_Analysis.

2.4.5.7 GUS Expression in Arabidopsis Seedlings

All *A. thaliana* were germinated and grown in Sungro Sunshine Mix #1 soil in a Percival growth chamber at 22 °C and 60% humidity with short day (8/16 h light/dark) or long day (16/8 h light/dark) conditions at a daytime PPFD of ~200 µmol/m2 s. *A. thaliana* stably expressing a GFP-GUS fusion protein driven by various PCONS promoters were generated by floral dipping of healthy Col-0 plants in PCONS binary vector-harboring GV3101 *A. tumefaciens* and subsequent seed selection with kanamycin on 1/2 strength Murashige and Skoog (MS) agar media. T2 seeds were germinated and GUS expression was visualized after 10 days. GUS expression was visualized using an adjusted method based on prior established protocols. (437) 10 day-old Arabidopsis seedlings were submerged in staining solution consisting of 0.5 mM K3Fe(CN)6, 0.5 mM K4Fe(CN)6, 100 mM sodium phosphate buffer pH 7.0, 10 mM ethylenediaminetetraacetic acid, and 0.5 mg/mL X-Gluc (5-bromo-4-chloro-3-indolyl-beta-d-glucuronic acid cyclohexylammonium salt) for 24 h at room temperature. Stained seedlings were sequentially washed with 70% and 100% ethanol for 8 h each. As a control, wild-type *A. thaliana* seedlings (Columbia ecotype) were studied using the same procedure.

2.4.5.8 Hairy Root Transformation

Medicago truncatula seeds were scarified in sulfuric acid and surface sterilized with 5% v/v commercial bleach with deionized water before being plated on 1/2 strength Murashige and Skoog (MS) agar media. After 2 days of stratification at 4 °C in the dark followed by two 2 of incubation at 22 °C, seedling radicles were removed with a scalpel and the wound inoculated with *Agrobacterium fabrum* ARqua1 harboring binary vectors containing PCONS promoters driving DsRed fluorescent protein expression. (438) Inoculated seedlings were co-cultured on 1/2 strength MS agar media for 5 days and then selected on ½ strength MS agar media with kanamycin (25 micrograms/mL) and 0.1% v/v Plant Preservative Mixture. Images were taken 3 weeks after inoculation.

2.4.5.9 Microscopy

Imaging of stable integration Arabidopsis GUS seedlings and Medicago DsRed seedlings was performed on an Olympus SZX16 stereoscope fitted with an Olympus SZX2-ILLT base for transmitted/oblique illumination and a fluorescence filter wheel with an X-Cite 120 excitation light source for fluorescence. Filters (Semrock BrightLine) for 562 nm (40 nm bandwidth) and 624 nm (40 nm bandwidth) were used for excitation of DsRed and emission filtering, respectively. Images were collected with a Lumenera INFINITY 2–5 color camera (gain = 20, exposure = 400 ms) with 0.63× adapter.

2.4.5.10 Data Availability

All barcode sequencing reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under the BioProject accession PRJNA922233 (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA922233/).

2.4.6 Guide to Supplementary Information Available Online

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.3c00075.

Supplementary Figure 2-9. Pairwise Pearson correlation of mRNA abundance between chosen samples to enrich for PCONS promoters

Supplementary Table 2-8. primer sequences

Supplementary Table 2-9. PCONS sequences

Supplementary Table 2-10. samples used for PCONS

Supplementary Table 2-11. plasmid information

2.4.7 Author Contributions

A.Z. and P.M.S. conceived and interpreted experiments and wrote the article; A.Z., L.D.K., I.J.O., L.J.W., N.F.C.H., C.W.G., S.N.T., and C.R.B. performed experiments; and A.Z. analyzed the data and prepared the figures; P.M.S. and H.V.S. provided supervision of research.

2.4.8 Acknowledgements

We thank Dr. Simon Alamos for input and feedback on the manuscript. This work was supported by the DOE Joint BioEnergy Institute, which is supported by the Office of Science, Office of Biological and Environmental Research, the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A.Z. was supported by NSF

Postdoctoral Fellowship in Biology, Award Number 2009093. Figures 2-9 and 2-12 were created in part with Biorender.com

2.5 Biosynthesis of Strained Amino Acids Through a PLP-Dependent Enzyme via Cryptic Halogenation

2.5.1 Summary and Personal Contribution

Sosa et al. 2024 is an example of the benefits in collaboration across disciplines. The bacteria *Pseudomonas azotoformans* was found to produce pazamine, a metabolite which shared structural similarities with the ethylene precursor 1-aminocyclopropane-1-carboxylate (ACC), via novel biochemical processes with potential application in chemical biosynthesis. I contributed the design, execution, data collection, and analysis informing the potential effect of pazamine on plant root-bacteria associations. Growth setups enabling continuous imaging of *Arabidopsis thaliana* root systems inoculated with a series of *P. azotoformans* exhibiting varied metabolic profiles were paired with software to process images into quantitative datasets, providing the ability to survey high sample numbers across multiple independent variables.

2.5.2 Text

α-Amino acids (AAs) serve as a diverse group of chiral building blocks used to construct a broad range of structures in either a templated or non-templated fashion by biological systems. A subset of twenty standard AAs is genetically encoded for ribosomal protein synthesis and includes aliphatic, aromatic, acidic, and basic sidechains. However, these proteinogenic AAs represent only a small fraction of the chemical diversity found in nature.(439) Indeed, both the AA monomers themselves and their resulting peptides can be highly modified to produce both new non-canonical AAs (ncAAs) and peptides of complex structure made either ribosomally (440) or non-ribosomally. (441) These downstream reactions include structural changes that affect backbone structure such as epimerization,(442) *N*-alkylation,(443) *N*-hydroxylation,(444) and cyclization.(445) Modification also occurs to incorporate new functionalities, such as hydroxyl groups (446), halogens,(447) alkenes and alkynes,(448) N-N bonded motifs like diazo,(449) hydrazine,(450) and diazeniumdiolate groups,(451) oxidized amines like hydroxylamines and nitro (452) groups, heterocycles like aziridines(453) and azetidines(454), or incorporation of unusual elements such as fluorine,(455) arsenic,(456) and selenium. (457) As such, AAs serve as an important source of diversity generation in biosynthesis, yielding the rich natural product structure found in ncAAs, alkaloids, nonribosomal peptides (NRPs), ribosomally-synthesized and post-translationally modified peptides (RiPPs) as well as new functionality in proteins found in post-translationally derived cofactors.(458,459)

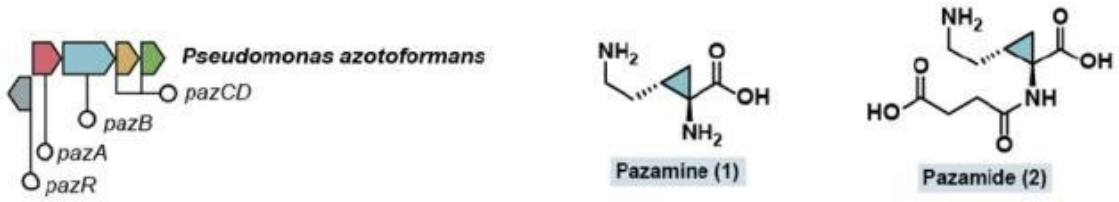
In particular, the study of ncAA biosynthesis has uncovered many interesting structural motifs and biosynthetic transformations.(439,448) After the identification of a new strategy for terminal alkyne formation in the ncAA, β-ethynylserine (βes), through cryptic chlorination(448) we became interested in exploring the role of other BesD halogenases in biosynthesis of AA-derived natural products. The BesD family was the first Fe(II)/α-ketoglutarate (αKG)-dependent radical halogenase family found to halogenate free AAs by activating C(*sp*3)-H bonds on the methylene backbone.(447) Bioinformatic analysis

shows that they are located in a wide range of biosynthetic contexts and we were therefore interested to explore the different outcomes.

In this work, we describe the discovery of the cyclopropane amino acid pazamine (**1**) and its derivative pazamide (**2**) (Figure 2-15A). Pazamine is made from lysine via a remarkably efficient two step pathway consisting of PazA and PazB, where radical chlorination by PazA enables a pyridoxal phosphate (PLP)-dependent cyclopropanation carried out by PazB. Studies with *Arabidopsis thaliana* seedlings suggest that the physiological function of **1** could be related to inhibition of ethylene biosynthesis. Furthermore, PazB can produce carbocycles of different ring sizes and be applied to the biocatalytic production of a cyclobutane amino acid.

Figure 2-15. Discovery of a biosynthetic cluster that produces a cyclopropane amino acid.

A



B

Gene	Swiss-Prot Homologue	%ID/%Pos	Proposed Function
pazR	HTH-type transcriptional regulator FL11(*Pyrococcus furiosus*)	32/51	Regulatory protein
pazA	L-lysine 4-chlorinase (*Streptomyces cattleya*)	56/71	Halogenase
pazB	Serine hydroxymethyltransferase (*Methanopyrus kandleri*)	40/59	Cyclopropanase
pazC	Cysteine/O-acetylserine efflux protein (*Salmonella typhimurium*)	29/47	Transporter
pazD	Cysteine/O-acetylserine efflux protein (*Salmonella typhimurium*)	28/41	Transporter

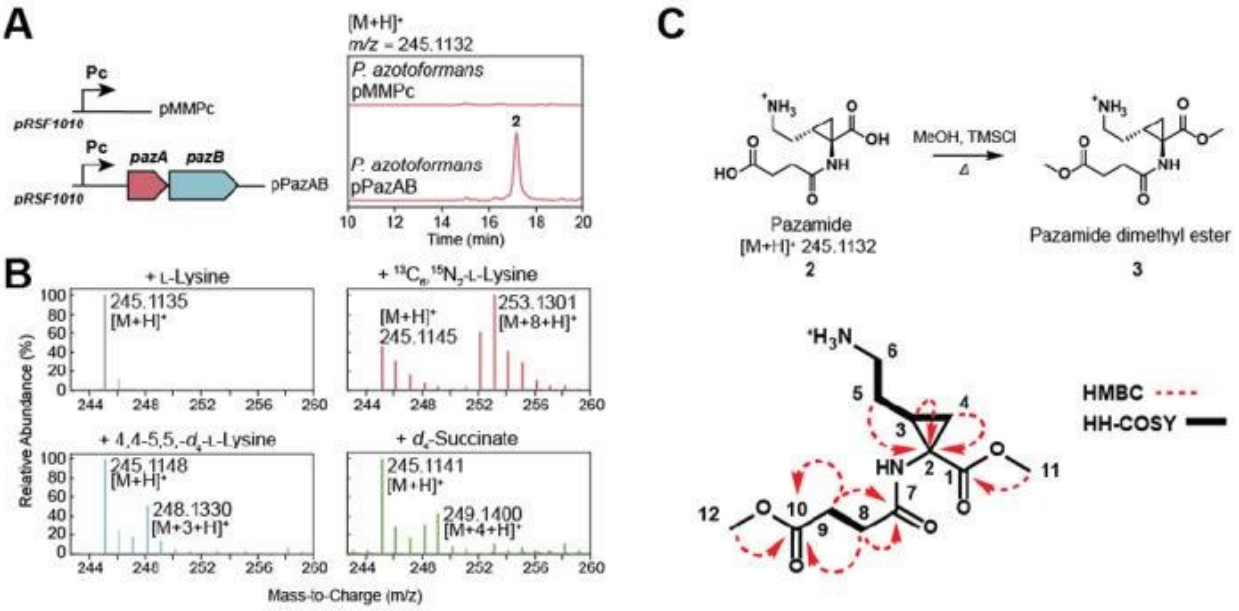
(A) A biosynthetic gene cluster required for the biosynthesis of cyclopropane-containing amino acid 1 and its downstream derivative 2 in P. azotoformans. **(B)** A bioinformatic analysis of the genes in the pazRABCD gene cluster. The genes are annotated based on Swiss-Prot homologues identified by a BLAST search and the percent identity (%ID) and percent similarity (%Pos) are shown. The proposed functions of the genes are included, based on the results of this study.

The *pazRABCD* cluster was first identified in *Pseudomonas azotoformans* through its BesD-like halogenase. The cluster encodes a regulatory protein (PazR), a lysine halogenase (PazA), a serine hydroxymethyltransferase-like enzyme (PazB), and two putative amino acid transporters (PazCD) (Figure 2-15B, Supplementary Table 2-12, Supplementary Figure 2-10). Serine hydroxymethyltransferases are a class of PLP-dependent enzymes which catalyze the interconversion of serine and glycine using tetrahydrofolate (THF) as a co-substrate, as well as folate-independent aldolase chemistry (Supplementary Figure 2-11).(460) A bioinformatic analysis revealed that PazB is distinct from the canonical bacterial SHMTs and a majority of its homologs are found in archaea (Supplementary Figures 2-10–2-11). Furthermore, residues involved in folate-dependent chemistry are not present in PazB, even though they are highly conserved in SHMTs from all domains of life (Supplementary Figure 2-11).(461,462)

Coupled with the fact that *pazB* is genetically colocalized with an amino acid halogenase, this information suggests that PazB may carry out a novel function related to the modification of (2*S*,4*R*)-chlorolysine produced by PazA. Additionally, this cluster and larger *pazAB*-containing clusters are found in other *Pseudomonas* spp. while select *Legionella* spp. encode a similar cluster containing a C4-lysine dichlorinase homolog (Supplementary Figure 2-12).

We set out to characterize the product of the *pazRABCD* gene cluster. Since PazB expresses very little soluble protein in *Escherichia coli* (Supplementary Table 2-13), these studies were carried out in the native host, *P. azotoformans*. The PazAB overexpression plasmid, pPazAB, was constructed by inserting the native *pazAB* fragment into the pMMPc-Gm plasmid under control of the Pc promoter from *Delftia acidovorans*.(463) The *P. azotoformans* pPazAB overexpression strain was then cultured in parallel with the *P. azotoformans* pMMPc-Gm empty plasmid control strain. The intracellular metabolome was extracted with acidic methanol and analyzed by high-resolution liquid chromatography-quadrupole time-of- light mass spectrometry (LC-QTOF). Differences in the metabolomes of these two strains were assessed using MS-DIAL(464) (Supplementary Figure 2-14), revealing a metabolite (*m/z* = 245.1132 [M+H]+) that is overexpressed in *P. azotoformans* pPazAB but absent in the control strain (Figure 2-16). Based on the calculated molecular formula (C10H16N2O5) and the absence of a 37Cl isotope pattern in the mass spectrum, it seemed possible that the biosynthesis of the PazAB product could involve cryptic chlorination. *Pseudomonas azotoformans* pPazAB was grown in the presence of the isotopically labeled precursors 4,4',5,5'-*d*4-L-lysine and 13C6,15N2-L-lysine to further validate this metabolite's origins. These studies showed that L-lysine is indeed incorporated into this product and that a deuteron from either C4 or C5 of lysine is removed, consistent with radical halogenation of lysine by PazA. In addition to lysine, there was another four-carbon fragment incorporated into the product. This fragment was determined to originate from succinate, which was confirmed by feeding experiments with *d*4-succinate (Figure 2-16B).

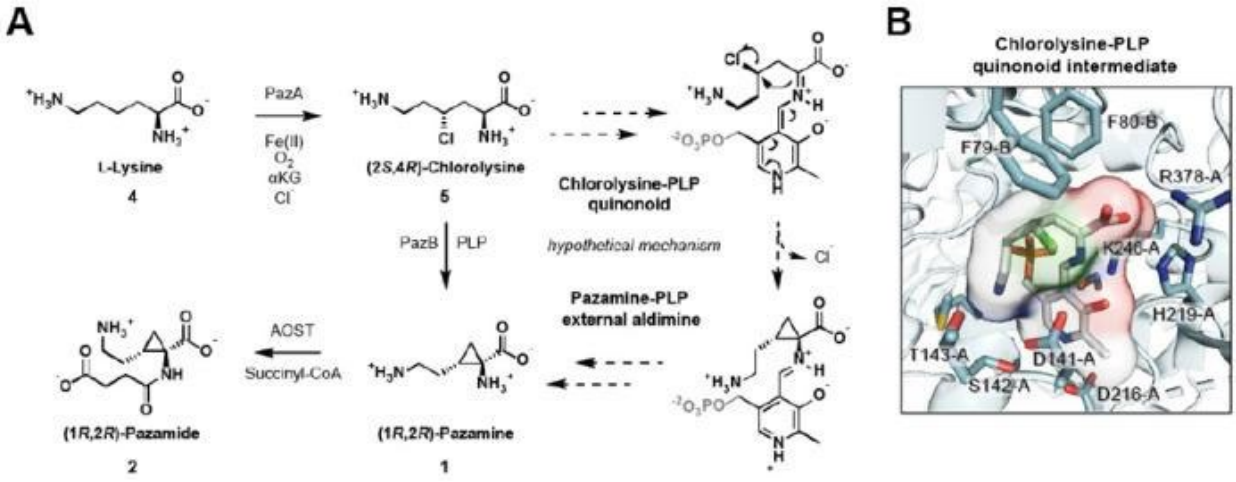
Figure 2-16: PazAB produces a cyclopropane amino acid.



(A) P. azotoformans transformed with either the PazAB expression plasmid (pPazAB) or the empty vector control (pMMPc) were cultured. Metabolomic analysis of the extracted cell culture revealed a new metabolite of m/z = 245.1132 [M+H]+ that was only detected when PazAB were expressed. **(B)** Isotopic labeling studies showed that the M+8 isotopologue is formed upon feeding of 13C6,15N2-l-lysine, confirming that the product is derived from l-lysine. Production of the M+3 isotopologue upon feeding of 4,4',5,5'-d4-l-lysine is consistent an H-abstraction step catalyzed by PazA during chlorination. Feeding of d4-succinate leads to production of the M + 4 isotopologue, indicating that succinate is the source of the additional four carbon fragment. **(C)** Compound 2 was isolated and methyl esterified for structural elucidation by 2D-NMR techniques. Key HMBC correlations are shown as red arrows while HH-COSY correlations are indicated with bold lines.

This compound was purified by preparative HPLC using hydrophilic interaction chromatography (HILIC). The product was analyzed by 2D-NMR following methyl esterification and determined to be a succinylated cyclopropane-containing amino acid (**3**) (Figure 2-16C, Supplementary Figure 2-14–2-16). The free amino acid parent compound **1** and its succinylated derivative **2** have been given the trivial names pazamine and pazamide, respectively, after the host organism *P. azotoformans*. With the product structure in hand, a biosynthetic hypothesis for **1** can be proposed where the PazA halogenase initiates the pathway by radical chlorination of the C4 of L-lysine (**4**) (Figure 2-17A). *In vitro* reactions of PazA with L-lysine confirm that it does indeed produce (2S,4R)-chlorolysine (**5**) (Supplementary Figure 2-17), which serves as a substrate for PazB. PazB is a PLP-dependent enzyme, suggesting that the carbocycle is made via deprotonation of Cα followed by nucleophilic attack on C4 with the chloride serving as a leaving group.

Figure 2-17: Proposed biosynthetic pathway to pazamine and pazamide.



(A) Cyclopropanation is proposed to take place in two steps. In the first step, radical halogenation of 4 by PazA incorporates a leaving group for the next step. In the second step, the PLP-dependent enzyme PazB reacts with 5 to form the external aldimine. Cα-deprotonation yields the quinonoid carbanion intermediate that can carry out an SN2-like reaction at C4 to form the carbocycle with stereoinversion and concomitant loss of chloride. 1 is then released from the PLP cofactor and is succinylated by AOST to produce 2. **(B)** A model of the PazB active site was generated by AlphaFold and the quinonoid intermediate before carbocyclization was docked in with AutoDock Vina. In this model, a polar pocket is observed and proposed to bind the ε-ammonium group in order to bring C4 into proximity of Cα for cyclization. The bending of the chain also appears to be assisted by steric interactions with F79 and F80.

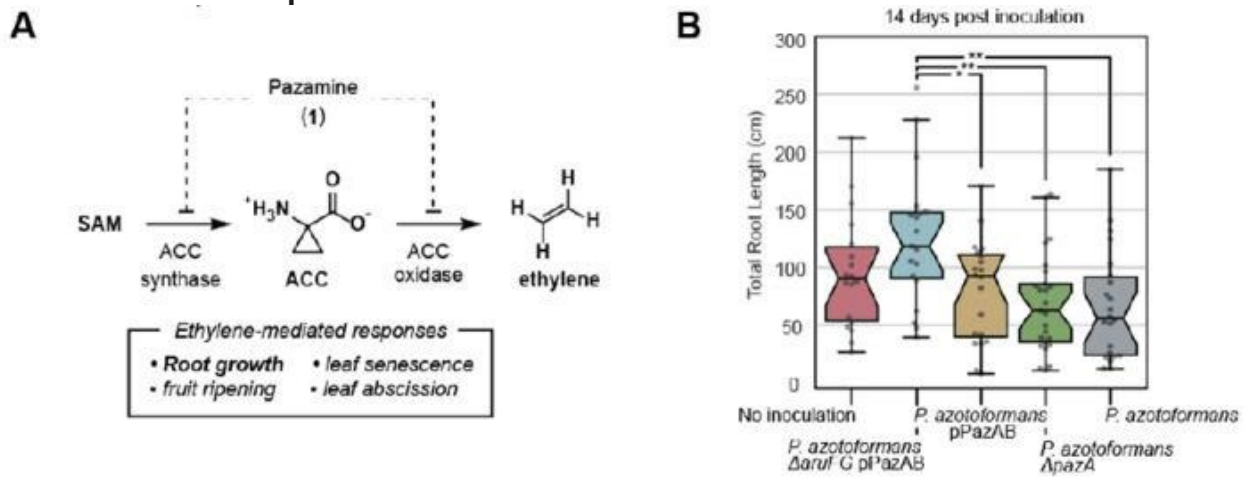
Several PazB orthologs were tested but found to be insoluble for heterologous expression in *E. coli.* (Supplementary Table 2-13). Thus, PazB was modeled using AlphaFold(465,466) in order to gain more insight into the cyclopropanation reaction (Supplementary Figure 2-18). AutoDock Vina(467,468) was then used to generate docked poses of the putative quinonoid intermediate of PLP and **5**. Comparison to the crystal structure of a canonical SHMT from *E. coli* bound to substrate(461) (PDB: 1DFO) assisted in identifying a biologically relevant pose (Figure 2-12B). Analysis of the model revealed the presence of a polar pocket in the PazB active site (D141-A, T142-A, and S143-A) that may interact with the Nε of PLP-bound **5** (Figures 2-12B and Supplementary Figure 2-19B). This interaction appears to facilitate folding of the methylene chain into a reactive conformation so that C4 can be brought into proximity of the Cα carbanion nucleophile. Additional steric interactions with F79-B and F80-B further appear to control positioning of the aminoalkyl chain into a productive conformation for cyclopropanation. Intramolecular attack by the Cα carbanion nucleophile on the chlorinated C4 results in chloride elimination and C-C bond formation with stereoinversion to yield the cyclopropane ring of **1**.

Given the absence of a gene candidate for a succinyltransferase in the *pazRABCD* operon or its genome neighborhood, we hypothesized that formation of **2** from **1** was not carried out by a dedicated enzyme but by an enzyme from primary metabolism. Examination of metabolic pathways involving the structurally-similar amino acid L-ornithine identified arginine/ornithine succinyltransferase (AOST), an enzyme which catalyzes the α-amino succinylation of both L-arginine and L-ornithine (Supplementary Figure 2-20).(469) To test this hypothesis, an AOST-knockout strain *P. azotoformans*

ΔaruFG was generated. Overexpression of PazAB in this strain shows that it still produces **1** but **2** is no longer observed (Supplementary Figure 2-21), which is consistent with a model where succinylation of **1** is carried out by AOST. Interestingly, AOST from *P. aeruginosa* PAO1 has been shown to be inhibited by and inactive towards D-ornithine, suggesting that **1** and L-ornithine have the same relative stereochemistry at Cα.(469) This information, the predicted stereoinversion of the cryptically chlorinated C4, and an NOE supported by modeling (Supplementary Figure 2-16) leads to our tentative assignment of the stereochemistry of **1** (and **2**) as (1*R* 2*R*).

As many Pseudomonads, including *P. azotoformans*, are known to associate with plant hosts,(470) we thought to question the bioactivity of **1**. We noticed that **1** bears a resemblance to another cyclopropane amino acid, 1-aminocyclopropane-1-carboxylate (ACC), an intermediate in the biosynthesis of ethylene in plants. This molecule is made through the PLP-catalyzed cyclization of the methionine moiety of *S*-adenosylmethionine (SAM), liberating methylthioadenosine in the process.(463,471) We thus hypothesized that **1** might interact with the ethylene pathway in a plant host, possibly as an inhibitor of either ACC synthase or ACC oxidase, lowering the amount of ethylene produced by plants (Figure 2-18A). Plants respond to the initial detection of bacteria by inducing ethylene production prior to downstream determination of bacterial lifestyle — i.e. type of pathogen or mutualist — and subsequent immune responses.(472) This results in an initial reduction of root length and increase in root branching and root hair formation unless the bacteria has processes that interfere with this response.(473,474) Inoculation of *Arabidopsis thaliana* with *P. azotoformans ΔaruFG* pPazAB demonstrated a rescuing of root length inhibition caused by bacterial inoculation and subsequent ethylene production. Wild type *P. azotoformans* and the **1**-deficient *P. azotoformans ΔpazA* are unable to rescue the reduction in root length. Interestingly, *P. azotoformans* pPazAB is also unable to fully rescue the observed root phenotype, possibly because it can sequester the proposed bioactive **1** as **2** (Figure 2-18B, Supplementary Figure 2-14). Overall, these data align with our understanding of the plant response to bacterial inoculation. When compared to an uninoculated control, the strain altered to increase **1** production induced the expected change in root architecture, without a reduction in root length. While these results support our initial hypothesis that **1** is capable of inhibiting ethylene biosynthesis, further work is required to definitively assess if and how this occurs.

Figure 2-18: Pazamine-producing bacteria rescue root length suppression in inoculated Arabidopsis thaliana.



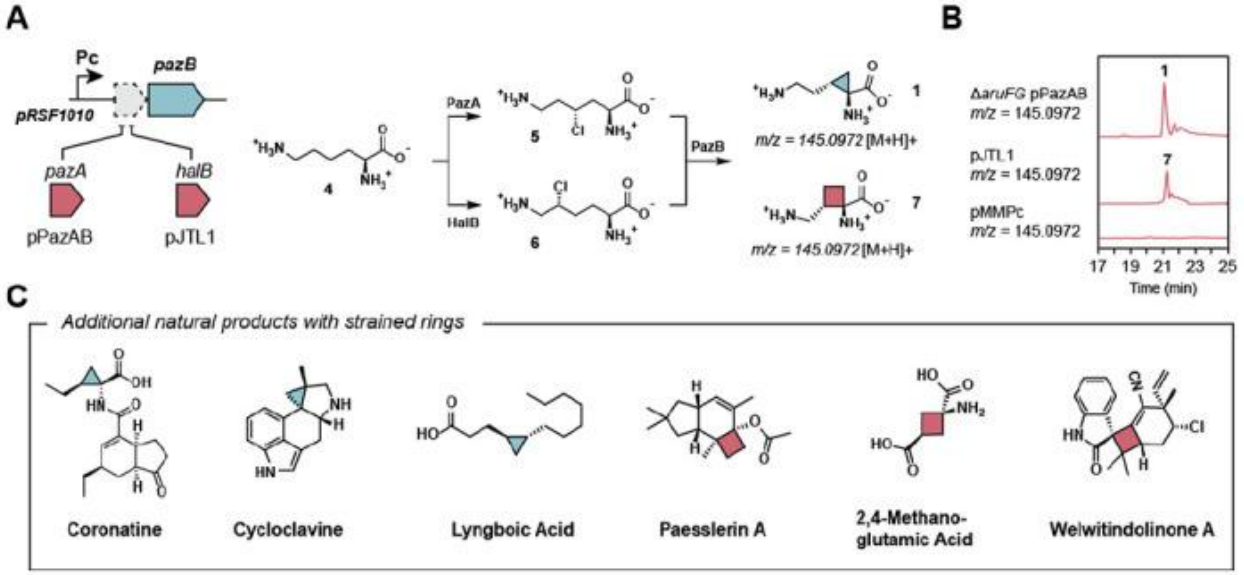
(A) As 1 is structurally similar to the ethylene precursor ACC, it may be capable of inhibiting enzymes involved in plant ethylene biosynthesis. **(B)** Inoculation of A. thaliana with P. azotoformans ΔaruFG pPazAB, a 1 producer, results in a rescue of the root length phenotype caused by bacterial inoculation and subsequent ethylene production. In P. azotoformans pPazAB, 1 can be converted to 2, diminishing the observed effect. This effect is statistically significantly (ANOVA and Tukey post-hoc test, * = p-value < 0.05, ** = p-value < 0.01) at 14 days post inoculation (DPI).

Strained carbocycles like the cyclopropane found in **1** are important structural elements found in a broad range of natural product structures.(475) Cyclopropanes themselves are metabolically stable motifs that are present in a number of pharmaceuticals based on their rigid, biplanar nature that can enforce desired conformations while providing unique hybridization in between *sp2* and *sp3*.(476) Along with cyclobutane rings, they are of high interest as substrates for C-C activation and reactivity towards nucleophiles and electrophiles due to their inherent ring strain.(477) In addition to synthetic applications, cyclopropane rings are important in nature in both natural products and lipids, where they they arise mostly via SAM-dependent methyl transfer to double bonds(478) as well as terpene cyclase-(479) and metalloenzyme-catalyzed rearrangements.(452,480) Another cyclopropane amino acid from *Pseudomonas* spp., coronamic acid, is made via cryptic chlorination of a carrier protein-bound amino acid followed cyclization by a Zn2+-dependent enzyme.(481) A handful of cyclobutane amino acids have also been found, such as 2,4-methanoproline and 2,4,-methanoglutamic acid(482) but their biosyntheses are not known. Other cyclobutane-containing natural products are thought to arise from terpene cyclase-catalyzed rearrangements of terpenes or by photochemical cycloadditions.(483)

Considering the scarcity of biochemical methods to generate cyclobutanes, we wanted to test if PazB could accept an alternative substrate to catalyze cyclobutanation. The enzyme HalB catalyzes the stereoselective halogenation of lysine at C5 so co-expression of HalB with PazB could enable the production of the cyclobutane amino acid **7** (Figure 2-19A). Towards this end, we made the plasmid pJTL1 to construct this pathway and we expressed it in *P. azotoformans*. Gratifyingly, we observed the production of a unique metabolite corresponding to **7** (*m/z* = 145.092 [M+H]+) (Figure 2-19B). Feeding 2,3,3',4,4',5,5',6,6'-*d*9-L-lysine to *P. azotoformans* pPazAB or pJTL1

showed that the expected M+7 isotopologues of **1** and **7** are produced, indicating that two deuterons are lost as would be consistent with radical halogenation and PazB-catalyzed cyclization (Supplementary Figure 2-21). These studies show that PazB is sufficiently promiscuous to expand ring size and accept substrates that are chlorinated at either C4 or C5. Furthermore, this system provides an avenue to study enzymatic cyclobutane synthesis, a rare and underrepresented biochemical reaction, as well as a potential biocatalytic route to various carbocyclic ncAAs.

Figure 2-19: Engineering the biosynthetic production of new carbocycle-containing amino acids.



(A) Combinatorial expression of PazB with different halogenases allows for the biosynthesis of various strained ncAAs. HalB halogenates l-lysine (4) at C5 altering the ring size of the PazB product. **(B)** P. azotoformans ΔaruFG pPazAB and P. azotoformans pJTL1 cultures produce 1 and 7, respectively. **(C)** Selected natural products which contain either a cyclopropane or a cyclobutane ring.

The discovery of the PazAB pathway provides an efficient enzymatic route towards formation of strained carbocyclic amino acids. The strategy of cryptic chlorination followed by PLP-catalyzed cyclization is a new variant for enzymatic cyclopropane formation that we have shown to be generalizable to production of new cyclopropane and cyclobutane amino acids. Specifically, PazB demonstrates catalytic plasticity in accepting different chlorinated amino acids for carbocycle formation, allowing ring size to be altered Analysis of the active site suggests simple electrostatic and steric interactions assist in folding the chain for cyclization, providing a design template for engineering other PLP-dependent enzymes for carbocycle formation. Indeed, the use of free amino acids by the PazAB pathway rather than tethered or specialized amino acids potentially allows greater scope for engineering new pathway variants or their incorporation into synthetic compounds or natural products utilizing amino acid building blocks. Taken together, the continual exploration of biosynthetic pathways can enable advances in biocatalysis and biosynthesis by the discovery of new reaction mechanisms and natural products for expanding the scope of enzymatic chemistry.

2.5.3 Materials and Methods

2.5.3.1 Commercial Materials

L-Lysine, iron(II) sulfate heptahydrate, ammonium formate (LC-MS grade), Dowex 50WX8 hydrogen form, irgasan (Irg) α-ketoglutarate (αKG) disodium salt dihydrate, lysozyme, 2-mercaptoethanol (βME), polyethyleneimine (PEI), pyridoxal 5′-phosphate (PLP), sodium dodecyl sulfate, sodium ascorbate, sucrose, trimethylchlorosilane (TMSCl) and were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile, agarose, bromophenol blue, carbenicillin disodium salt, diethyl ether, dithiothreitol (DTT), deoxynucleotides (dNTPs), ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA), ethyl acetate, formic acid, hydrochloric acid, gentamicin sulfate (Gm), HEPES, imidazole, kanamycin sulfate (Km), methanol, GeneRuler 1 kb Plus DNA Ladder, PageRuler Plus Prestained Protein Ladder, sodium chloride, sodium hydroxide, were purchased from Thermo Fisher Scientific (Waltham, MA). Acetic acid, dimethylsulfoxide, glycerol, LB Miller Agar, LB Miller Broth, nutrient broth, magnesium chloride hexahydrate, Terrific Broth, were purchased from EMD-Millipore (Burlington, MA). Absolute ethanol was from VWR International (Radnor, PA). InstantBlue Protein Stain was from Expedeon (San Diego, CA). Isopropyl ß-D-1-thiogalactopyranoside (IPTG) was from Santa Cruz Biotechnology (Dallas, TX). Bradford assay reagent concentrate and ethidium bromide were purchased from Bio-Rad Laboratories (Hercules, CA). Deuterium oxide, L-lysine dihydrochloride ($^{13}C_6$, 99%; $^{15}N_2$, 99%), L-lysine dihydrochloride (4,4′,5,5′-d_4), L-lysine dihydrochloride (2,3,3′,4,4′,5,5′,6,6′-d_9), and disodium succinate (2,2′,3,3′-d_4) were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Phusion polymerase, Phusion HF buffer, all restriction enzymes, restriction enzyme buffer (CutSmart), and Taq ligase were from New England Biolabs (Ipswitch, MA). T5 exonuclease was from Epicentre (Madison, WI). Ni-NTA agarose resin and DNA purification kits were purchased from Qiagen (Redwood City, CA). Oligonucleotides and gBlocks gene fragments were synthesized by Integrated DNA Technologies (Coralville, IA) or Twist Bioscience (South San Francisco, CA). All chemicals were used as purchased without further purification.

2.5.3.2 Bacterial Strains and Culture

Escherichia coli DH10B-T1^R was used for cloning, *E. coli* BL21 Star (DE3) was used for protein expression and purification, and *E. coli* SM10 was used to transfer *oriT*-containing plasmids via bacterial conjugation. Unless otherwise stated, *E. coli* liquid cultures were grown at 37°C and 200 RPM on LB Miller broth with appropriate antibiotics, while single colonies were grown on LB Miller agar plates (1.5% *w/v*) at 37°C. *Pseudomonas azotoformans* (DSM 18862) was purchased from the German Collection of Microorganisms and Cell Cultures (DSMZ). *Pseudomonas azotoformans* strains were grown in LB Miller broth at 30°C and 200 RPM with appropriate antibiotics, while single colonies were grown on LB Miller agar plates (1.5% *w/v*) at 30°C. *P. azotoformans* strains were stored as glycerol stocks at -80°C and streaked to single colonies before cell culture for various experiments

.

2.5.3.3 Bacterial Transformation

To 100 µL of chemically competent *E. coli*, plasmid DNA, 20 µL of KCM solution, and water up to a final volume of 200 µL were added. The mixture was incubated on ice for 30 min prior and then incubated at 42°C for 90 s. The cells were returned to ice for 2 min,

600 uL of LB Miller broth was added, and the cells were recovered at 37°C for 1 h before plating on selective media. *Pseudomonas azotoformans* was transformed via electroporation. A freezer stock of *P. azotoformans* was streaked out onto LB Miller agar and incubated overnight at 30°C. A single colony was used to inoculate 5 mL of LB Miller broth and the starter culture was grown at 30°C, 200 RPM overnight. The following morning, 25 mL of LB Miller broth was inoculated with 1% (v/v) starter culture and grown at 30°C, 200 RPM until OD_{600} of 0.4 – 0.6. The culture was chilled in an ice bath for 15 min and then centrifuged at 5000 × *g* for 10 min. The supernatant was gently decanted, and the cell pellet was washed with 10 mL of chilled electroporation solution (1 mM HEPES, 1 mM $MgCl_2$, pH 7.2). The cells were once again pelleted at 5000 × *g* for 10 min and then resuspended in 1 mL of electroporation solution. The cells were washed as such two more times before resuspension in 500 µL of electroporation solution. Then, 100 µL of electrocompetent *P. azotoformans* was added to a chilled 2 mm gap width electroporation cuvette along with 50 - 100 ng of plasmid DNA. The cells were electroporated with a Bio-Rad MicroPulser on the Ec2 setting (2.5 kV, 5 ms). The cells were immediately resuspended in 800 µL of LB Miller broth and recovered at 30°C, 200 RPM for 90 min prior to plating on selective media and incubating at 30°C overnight.

2.5.3.4 Bacterial Conjugation

Plasmid DNA was also introduced to *P. azotoformans* by conjugation. First, *E. coli* SM10 was transformed with the plasmid to be conjugated by the heat shock method described above. Then, a freezer stock of *P. azotoformans* was streaked out onto LB Miller agar and incubated overnight at 30°C. A single colony of *E. coli* SM10 and *P. azotoformans* were separately used to inoculate 5 mL of LB Miller broth. When the cultures each reached an OD_{600} of 0.8-1.0, they were mixed 1:1 to a final volume of 1 mL. The combined cells were then collected by centrifugation at 18,000 × *g* for 1 min and the supernatant was removed. The cell pellet was resuspended in 200 µL of LB Miller broth and plated directly in the center of an LB Miller agar plate and incubated overnight at 30°C. The next d, the bacterial mating mixture was scraped from the plate, resuspended in 1 mL of LB Miller. The resuspended mating mixture was diluted 1:10 and 1:100 in LB Miller broth and plated on LB Miller agar plates containing Gm (30 µg mL^{-1}) and Irg (20 µg mL^{-1}). After incubating overnight at 30°C, only transformed *P. azotoformans* remained.

2.5.3.5 Plasmid Construction

Gibson assembly and cloning by homologous recombination (HR) were routinely used to construct plasmids, using *E. coli* DH10B-$T1^R$ as the cloning host. Briefly, plasmids were first designed *in silico* via Benchling (Benchling, San Francisco, CA). The plasmid backbones were digested with the selected restriction enzymes and purified via the Qiagen PCR Cleanup Kit (Qiagen, Redwood City, CA) or amplified by Q5 polymerase (New England Biolabs, Ipswitch, MA). The insert amplicons were generated with either Phusion polymerase or GoTaq polymerase via the primers listed in *Table S1* and purified as above. For Gibson assembly, the inserts and backbone were mixed in 1:1, 1:3, and 1:5 ratios in 5 µL volumes, mixed with 15 µL of Gibson master mix (484), and incubated at 50°C for 1 h. The Gibson mix was transformed into chemically competent *E. coli* DH10B-$T1^R$ by heat shock, plated on the appropriate selective media, and incubated at 37°C overnight. For HR cloning, 100 ng of plasmid was mixed in a 1:2 molar ratio with

insert and directly transformed into *E. coli* DH10B-T1^{R} by heat shock, plated on the appropriate selective media, and incubated at 37°C overnight. Positive colonies were identified by colony PCR. Plasmids were sequenced at Genewiz to confirm the identity of the inserted sequence (Genewiz, South Plainfield, NJ).

2.5.3.6 Bioinformatic Analysis of the Pazamine Gene Cluster

The genes in the *pazRABCD* cluster were searched against the SwissProt database using BLAST (485) to identify similar proteins which have been characterized in the literature. The *pazB* gene was further investigated with the EFI-EST (486) web tools by generating a sequence similarity network (SSN) of the SHMT Pfam (PF00464) and PazB (A0A1V2JN15). Only sequences from bacteria and archaea were included. The SSN was further manipulated within Cytoscape (487) to generate additional SSNs.

2.5.3.7 Generating *P. azotoformans* Knockout Strains

Genes were disrupted in *P. azotoformans* following the protocol from Hmelo et al (488). Briefly, upstream and downstream genomic regions of the genes of interest were amplified from bacterial colonies of *P. azotoformans*. For every kilobase pair (kb) of DNA to be knocked out (*n*), the genomic homology arms of approximately 500 + *n**50 bp were amplified by PCR. Homology arms were designed to generate an in-frame deletion of the genomic region of interest, resulting in a small peptide product. Primers for these amplicons can be found in *Table S1*. These homology arms were cloned into the pEXG2 cloning vector by Gibson assembly as described above and were delivered to *P. azotoformans* by bacterial conjugation, as described above. A single colony of conjugated *P. azotoformans,* which has undergone a single crossover event, was picked and streaked onto a nutrient broth plate containing 1.5% (*w/v*) agar and 5% (*w/v*) sucrose and incubated at 30°C overnight. The following d, single colonies were screened by colony PCR for individuals which have successfully undergone a double crossover event and have lost the gene of interest.

2.5.3.8 HILIC/QTOF-MS Analysis of Metabolites

Samples were analyzed via hydrophilic interaction liquid chromatography (HILIC) coupled with high-resolution quadrupole time-of-flight mass spectrometry (QTOF-MS). Samples were analyzed using an Agilent 1290 UPLC (Santa Clara, CA) on a SeQuant ZIC-pHILIC (5 µm, 2.1 × 100 mm; EMD-Millipore, room temperature), with buffer A (90% acetonitrile, 10% water and 10 mM ammonium formate) and buffer B (90% water, 10% acetonitrile and 10 mM ammonium formate) comprising the mobile phase. The column was kept at 40°C. A linear gradient from 100% to 60% buffer A over 17 min, followed by a linear gradient from 60% to 40% buffer A over 8 min was performed at a flow rate of 0.2 mL min^{-1}. Mass spectra were acquired in positive-ionization mode using an Agilent 6530C QTOF mass spectrometer (gas temperature = 300°C, nebulizer = 35 psi, capillary voltage = 3500 V, nozzle voltage 500 = V, fragmentor voltage = 175 V, skimmer = 65 V, Oct 1 RF Vpp = 750 V).

2.5.3.9 Comparative Metabolomics

Starting from single colonies isolated from glycerol stocks, *Pseudomonas azotoformans* pPazAB and *P. azotoformans* pMMPc-Gm were grown in triplicate, as described above, in LB Miller broth (25 mL) containing Gm (30 µg mL^{-1}) in 250 mL baffled shake flasks.

After 24 and 72 h, 2 mL of cells were removed from the culture and centrifuged at 21,300 × *g* for 1 min. The supernatant was separated from the cell pellet and mixed 1:1 with 1% (*v/v*) formic acid in methanol. The wet weight of the cell pellet was measured and resuspended in 1% formic acid in methanol (5 µL μg^{-1} of wet cell pellet). The methanolic samples were vortexed to homogeneity and then centrifuged at 21,300 × *g* for 10 min. The supernatants were sampled without disturbing the pellet and analyzed by HILIC/QTOF-MS as described. The resulting data files (.d) were exported and converted to .abf files (Reifycs Analysis Base File Converter; https://www.reifycs.com/abfconverter/) and the resulting files were imported into and analyzed with MS-DIAL(464), a software package for untargeted metabolomics analysis For peak identification, an MS1 tolerance of 0.01 Da was used, along with a minimum peak height amplitude of 5000. For alignment, a retention time tolerance of 0.15 min and a MS1 tolerance of 0.01 Da was used. For chromatographic features to be included in the same group for analysis, >50% of the samples had to contain the chromatographic feature of interest (%N detected in at least one group). In the resulting alignment file generated by MS-DIAL, data in the ion table were restricted to *m/z* values between 100 and 300 and then sorted by fold-increase to identify differentially biosynthesized metabolites which were only present in the pPazAB overexpression condition. Metabolites with both high fold-increases and large signal-to-noise ratios (>20) were recorded as molecules of interest.

2.5.3.10 Stable Isotope Labeling

Starting from a single colony isolated from a glycerol stocks, *Pseudomonas azotoformans* pPazAB was grown in LB Broth (25 mL) containing Gm (30 µg mL^{-1}) in 250 mL baffled shake flasks as described above. After reaching OD_{600} of 0.8-1.0, 5 mL of the culture was transferred to a 50 mL glass culture tube and fed 20 mg of isotopically labeled substrate ($^{13}C_6$,$^{15}N_2$-L-lysine, 4,4´,5,5´-d_4-L-lysine, 2,3,3´,4,4´,5,5´,6,6´-d_9-L-lysine, 2,2´,3,3´-d_4-succinate) to a final concentration of 4 mg mL^{-1}. After another 24 h of growth, the metabolome was extracted from the cell pellet as described and analyzed by the general HILIC/QTOF-MS method.

2.5.3.11 Amino Acid Fmoc-Derivatization and QQQ-MS/MS Analysis of Metabolites

Unlabelled and isotopically labeled metabolomic samples were grown as described above. The cellular metabolome was extracted as described above except using pure methanol, without formic acid. To 50 µL of the metabolomic extract, we added 12.5 µL of 200 mM, pH 8 sodium borate buffer and 10 µL of 10 mM Fmoc chloride in dichloromethane. The mixture was briefly vortexed and allowed to react at room temperature for 10 min. The sample was then analyzed via reverse phase chromatography coupled with triple-quadrupole tandem mass spectrometry (LC-QQQ-MS/MS). Samples were analyzed using an Agilent 1290 UPLC (Santa Clara, CA) equipped with an Agilent Poroshell 120 EC-C18 column (2.7 µm, 2.1 × 50 mm). Solvent A (0.1% formic acid in water) and solvent B (acetonitrile) comprised the mobile phases. Solvent A was kept at 100% for 1 min, followed by a linear gradient from 0% solvent B to 100% solvent B for 5 min. The mobile phase was then held at 100% solvent B for 1 min, before returning to 100% solvent A over the next min. The flow rate was 0.6 mL min--$^{-1}$. Mass spectra were acquired in positive mode on an Agilent 6460 QQQ MS (300°C, gas flow = 5 L min^{-1}, nebulizer = 45 psi, capillary voltage = 3500 V, nozzle voltage = 500 V

fragmentor = 135 V, the collision energy = 10, cell accelerator voltage = 7). The product ions (*m/z* = 152.1 $[M+H]^+$) were detected in MRM mode from precursor ions resulting from Fmoc-derivatized, perdeuterated compounds 1 and 7 (*m/z* = 374.2 $[M+H]^+$).

2.5.3.12 Density functional theory (DFT) calculations

All DFT calculations were performed in Gaussian 16, Revision B.01 (489). Geometry optimizations were carried out using the hybrid three-parameter Becke´s (B3LYP) functional (490) with the empirical correction to dispersion (+GD3) and the def2-SVP basis set, with water as the solvent in the Self-Consistent Reaction Field (SCRF). Frequency calculations on the optimized geometries were performed in the same level of theory (B3LYP/GD3/def2-SVP/SCRF=water) to obtain the enthalpy and entropy corrections for the calculated energies.

2.5.3.13 NMR spectroscopy

All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe and was controlled using Topspin (Version 3.2) software. Data were processed using Topspin (Version 3.2) by zero-filling once in each dimension, followed by apodization, Fourier transformation, and phasing. Spectra were analyzed using Mnova software (Mestrelab Research, Escondido, CA, USA).

2.5.3.14 Isolation of Pazamide (2)

A glycerol stock of *P. azotoformans* pPazAB was streaked onto an LB Gm plate and incubated at 30°C for 36 h. A single colony was used to inoculate 50 mL of LB Gm media and an overnight culture was grown at 30°C, 200 RPM. The overnight culture was used to inoculate 6 × 500 mL (3 L) of LB media in 2.5 L baffled shake flasks and grown for 24 h at 30°C, 200 RPM. The cell pellets were harvested by centrifuging at 8000 × *g* for 5 min and extracted with 2 × 50 mL of 80% (*v/v*) methanol in water with 1% (*v/v*) formic acid. The pellets were resuspended to homogeneity by vortexing. The extract was centrifuged at 8,000 × *g* for 15 min and the supernatant was collected and dried by rotary evaporation. The extract was purified by preparative HPLC using an Eclipse XDB-C18 column (5 µm, 9.4 mm × 250 mm, Agilent) on an Agilent 1260 HPLC with 100% water run isocratically as the mobile phase (5 mL min^{-1}) for 15 min. Fractions containing pazamide were identified by LC-MS and were pooled and dried to give a semi-pure extract. The semi-pure extract was resuspended in 50% ACN/50% H_2O and further purified by preparative HPLC using a ZIC-HILIC column (5 µm, 21.1 mm × 50 mm, Supelco) on an Agilent 1200 HPLC with buffer A (10% ACN in water, 10 mM ammonium formate) and buffer B (90% ACN in water, 10 mM ammonium formate) comprising the mobile phase (5 mL min^{-1}). Buffer A was increased linearly from 10% to 40% at 8.08 min, 40% to 60% at 12.11 min, 60% to 70% at 18.17 min, and 70% to 10% at 20.19 min. Fractions containing 2 were identified by LC-MS, pooled, and evaporated to yield 22 mg of pazamide as an off-white powder. ^{1}H NMR (900 MHz, D2O) δ 3.08 (qt, J = 13.0, 6.5 Hz, 2H), 2.56 – 2.43 (m, 4H), 2.07 (dq, J = 11.1, 5.3 Hz, 1H), 1.95 (dq, J = 15.3, 7.8 Hz, 1H), 1.37 – 1.31 (m, 2H), 1.22 – 1.17 (m, 1H). ^{13}C NMR (226 MHz, D2O) δ = 176.61, 170.61, 39.59, 38.88, 31.48, 31.44, 26.32, 24.50, 20.32. Found QTOF/MS *m/z* = 245.1130 $[M+H]^+$ (Calculated *m/z* for $C_{10}H_{17}N_2O_5$ = 245.1132, -0.81 ppm error).

2.5.3.15 Pazamide Dimethyl Ester (3)

Pazamide 2 was methyl esterified (491) to facilitate structural elucidation. The product was stirred in acidic methanol prior to purification with HPLC using a ZIC-HILIC column (5 µm, 21.1 mm × 50 mm, Supelco) on an Agilent 1200 HPLC with buffer A (10% ACN in water, 10 mM ammonium formate) and buffer B (90% ACN in water, 10 mM ammonium formate) comprising the mobile phase. Buffer A was increased linearly from 10% to 40% at 8.08 min, 40% to 60% at 12.11 min, 60% to 70% at 18.17 min, and 70% to 10% at 20.19 min. Fractions containing 3 were identified by LC-MS, pooled, and evaporated to dryness via rotary evaporation and high vacuum. ^{1}H NMR (900 MHz, D2O) δ 3.72 (s, 3H), 3.68 (s, 3H), 3.06 (td, J = 7.1, 2.9 Hz, 2H), 2.66 (t, J = 6.7 Hz, 2H), 2.59 – 2.51 (m, 2H), 2.05 (dp, J = 12.8, 6.5 Hz, 1H), 1.95 (tt, J = 15.0, 7.2 Hz, 1H), 1.59 (dtd, J = 9.6, 8.4, 6.1 Hz, 1H), 1.54 (dd, J = 8.3, 5.3 Hz, 1H), 1.42 (dd, J = 9.6, 5.3 Hz, 1H). 13C NMR (226 MHz, D2O) δ 179.14, 178.13, 175.60, 55.73, 54.94, 41.60, 39.61, 32.56, 31.60, 30.83, 26.82, 24.69.

2.5.3.16 PazB Modeling and Substrate Docking

The protein sequence of PazB (UniProt: A0A1V2JN15) was supplied to the AlphaFold 2 Colab server, ColabFold(466), to generate a structural prediction of a PazB homodimer using the default parameters. Substrate models were generated from SMILES strings at the CACTUS server (http://cactus.nci.nih.gov/translate/). Substrate and protein models were prepared for docking and used in docking experiments with AutoDock Vina using the default parameters (467,468). Generated poses were compared to a substrate-bound *E. coli* GlyA (PDB: 1DFO) crystal structure (461) to identify catalytically relevant poses before further analysis.

2.5.3.17 Germination of *Arabidopsis thaliana*

Arabidopsis thaliana Columbia 0 (Col-0) ecotype seeds were sterilized with 20% (*v/v*) bleach for 10 min before washing five times with sterilized water. The seeds were resuspended in 0.1% (*w/v*) agarose to allow for precise placement on 1/3 Murashige and Skoog, 1% (*w/v*) plant tissue culture agar plates prior to germination. After drying the plates were sealed with micropore tape, wrapped in foil, and cold stratified at 4℃ for 48 h. The plates were then unwrapped and placed in a growth chamber with a photoperiod of 10 h light (intensity of 80 µmol m^{-2} s^{-1}), 14 h darkness, and set to 22℃ for germination. The relative humidity was maintained at 60%.

2.5.3.18 Plant Inoculation Assays

The same plates containing the germinated *A. thaliana* seeds were transferred to a growth chamber and grown for 5-7 d at 22℃ with a photoperiod of 16 h of light (intensity 130 µmol m^{-2} s^{-1}) and a dark period of 8 h. The relative humidity was maintained at 60%. The seedlings were then inoculated with overnight liquid cultures of *P. azotoformans* that were adjusted with LB Miller broth to an OD_{600} of 0.02. In a biosafety chamber, the plates were unsealed and the diluted bacterial culture (100 µL) was added to each individual *A. thaliana* plate in a line approximately 1 cm below the initial seedling placement. A negative control condition was prepared with nothing added. Plates were briefly left to dry before resealing with micropore tape and returning to the growth chamber. Plates were imaged

on the d of inoculation and at 2 d intervals until 14 d post inoculation (dpi) using an EPSON V600 photo scanner with a black background for contrast. Images were then imported to RhizoVision (199) where measurements were taken according to the same process and settings across the experiment. Data frames were analyzed in Jupyter notebook. One-way and two-way ANOVA analysis with Tukey´s post hoc test were used for testing differences in treatments. Statistical significance is denoted as * (p-value < 0.05) or ** (p-value < 0.01)**.**

2.5.3.19 Protein Expression and Purification

E. coli BL21 Star (DE3) cells were transformed, as described above, with the appropriate protein-expression plasmid. A single colony was used to inoculate an overnight culture of LB Miller broth and grown at 37°C, 200 RPM overnight. The next d, TB supplemented with the appropriate antibiotic was inoculated with 1% of the overnight culture and grown at 37°C, 200 RPM until the culture reached an OD_{600} of between 0.8 and 1.0. At this point, the cells were cooled on ice for 20 min before being induced with IPTG (100 µM). The proteins were then expressed overnight (12 -16 h) at 16 °C, 200 rpm. Cell pellets were collected by centrifugation at 8,000 × *g* for 5 min at 4 °C. The supernatant was decanted, and cells were resuspended in 5 mL g^{-1} of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 10 mM βME, pH 7.5) supplemented with 1 tablet per liter of cell culture of EDTA-free protease inhibitor cocktail (Roche). The resulting cell suspension was then lysed via sonication, using a Qsonica Q700 sonicator (Amplitude = 50, 5 s on, 25 s off, 2.5 min total process time, ½" tip). The lysate was then centrifuged at 13,500 × *g* for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (wt/vol) polyethyleneimine at 4 °C for 30 min, with periodic inversion. The precipitated DNA was then removed by centrifugation at 13,500g for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL g^{-1} of cell paste) for 45 min at 4°C, resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 10 mM βME, pH 7.5) for 15–20 column volumes. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 10 mM βME, pH 7.5). Fractions containing the target protein were pooled according to absorbance at 280 nm and concentrated using an Amicon Ultra spin concentrator (10-kDa MWCO; Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 10% (vol/vol) glycerol and 10 mM βME, pH 7.5) using PD-10 desalting columns. For PazB and PazB-like enzymes, 20 µM PLP was included in the lysis and wash buffers. Final protein concentration was estimated according to the absorbance at 280 nm. For the PazB and PazB-like enzymes, those which were insoluble were also co-expressed with the chaperones DnaK, DnaJ, and GrpE. In this case, the plasmid pKJE7 (Takara Bio Inc.) was co-transformed along with the PazB construct. The expression medium also contained 1 mg mL^{-1} L-arabinose. The procedure was otherwise unchanged.

2.5.3.20 *In Vitro* Assays for PazB Activity

All reactions (100 µL) contained L-lysine · HCl (2 mM), disodium αKG (6 mM), sodium ascorbate (2 mM), $Fe(SO_4)_2 \cdot 7H2O$ (100 µM), PLP (50 µM) and sodium chloride (> 20 mM, from protein storage buffer) in 50 mM HEPES buffer, pH 7.5. Reactions were initiated by addition of purified PazA (5 µM) and allowed to proceed for 45 min at room

temperature before addition of PazB variants to a final concentration of 10 μM. Additionally, PazB activity was also assayed in the presence of tetrahydrofolic acid (225 μM), divalent metals (Zn^{+2}, Mn^{+2}, Co^{+2} – 1 mM), or by starting 4-chlorolysine standard (2 mM). Upon addition of a PazB variant, the reaction was allowed to proceed overnight at room temperature. The reaction was then quenched with 5 volumes of methanol, 1% formic acid and centrifuged at 21,300 × *g* for 10 min at 4°C to remove the protein from the solution. The sample was then analyzed with the HILIC/QTOF-MS analysis described above.

2.5.4 Author Contributions

M. B. Sosa was responsible for designing experiments, performing experiments, collecting data, analyzing data, and writing the manuscript. J. T. Leeman was responsible for designing experiments, performing experiments, and collecting data. L. J. Washington was responsible for designing experiments, performing experiments, collecting data, and analyzing data. H. V. Scheller was responsible for designing experiments and validating data. M. C. Y. Chang was responsible for administering the project, designing experiments, validating data, and writing the manuscript. All authors were involved in editing the manuscript.

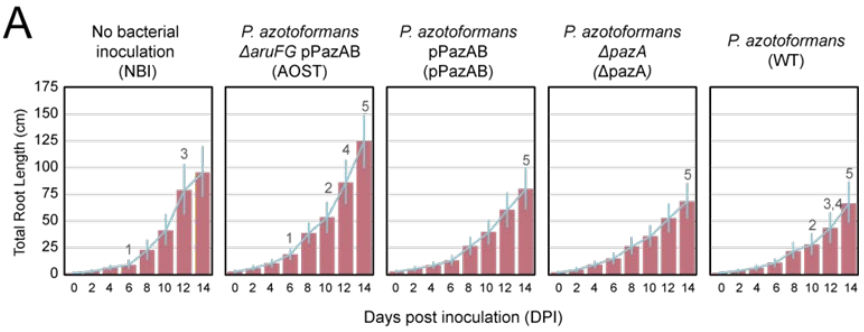
2.5.5 Acknowledgements

This work was funded by generous support from the NIH (R01GM134271). M.B.S. acknowledges the support of an NIH NRSA Training Grant (1 T32 GMO66698). L.J.W. acknowledges support from the NSF Graduate Research Fellowship Program (DGE2146752). L.J.W. and H.V.S. were funded by The Novo Nordisk Foundation grant no. NNF19SA0059362 (InRoot) and by the Joint BioEnergy Institute (http://www.jbei.org) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. We would also like to thank Dr. Ioannis Kiporous for generating DFT models of pazamide and Dr. Edward Koleski for generating the P.azotoformans ΔpazA knockout strain. Instruments in the UC Berkeley College of Chemistry NMR Facility were supported in part by NIH S10OD024998. Support for the 900 MHz NMR spectrometer in the QB3 Institute in Stanley Hall at University of California, Berkeley, was kindly provided by the NIH (GM68933).

2.5.6 Guide to Supplementary Information Available Online

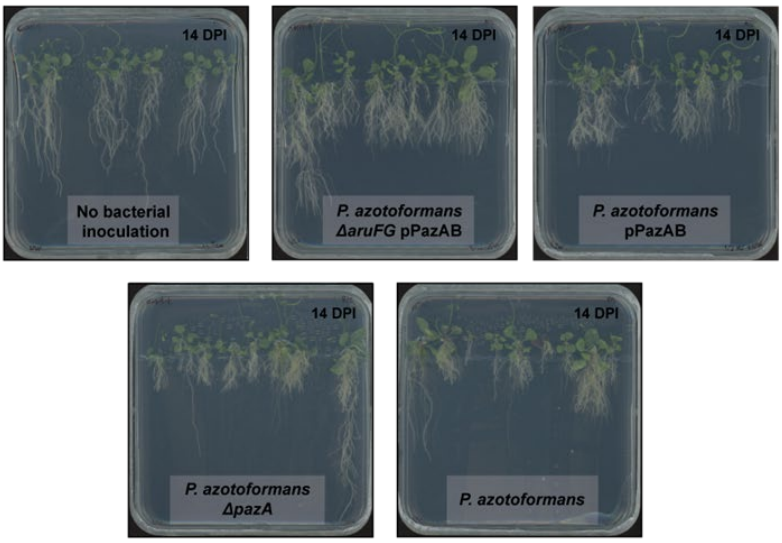
Supplementary Figure 2-10. SHMT and PazB sequence similarity network
Supplementary Figure 2-11. Serine hydroxymethyltransferase (SHMT) sequence analysis
Supplementary Figure 2-12. Gene neighborhoods surrounding the pazRABC cluster
Supplementary Figure 2-13. Metabolomic analysis with MS-DIAL
Supplementary Figure 2-14. 1D-NMR characterization of pazamide dimethyl ester (3)
Supplementary Figure 2-15. 2D-NMR characterization of pazamide dimethyl ester (3)
Supplementary Figure 2-16. 1D-NMR spectrum of pazamide (2)
Supplementary Figure 2-17. Confirmation of the PazA product
Supplementary Figure 2-18. Structural and sequence analysis of PazB active site
Supplementary Figure 2-19. Arginine/ornithine succinyltransferase can succinylate pazamine
Supplementary Figure 2-20. Arabidopsis seedling growth phenotype after bacterial inoculation

A



Annotation	Group 1	Group 2	*p*-value
1	ΔAOST	NBI	0.0374
2	ΔAOST	WT	0.0255
3	NBI	WT	0.0499
4	ΔAOST	WT	0.0087
5	ΔAOST	pPazAB	0.0378
5	ΔAOST	ΔpazA	0.0021
5	ΔAOST	WT	0.002

B



Arabidopsis thaliana

Arabidopsis seedling growth phenotype after bacterial inoculation. **(A)** The quantification of the total root lengths of inoculated seedlings over 14 d. Bars denote the mean, and the line plot denotes change over time and the mean (± s.d.). The numbers indicate statistically different means between groupings, described below the graph (*n*: NBI = 18, ΔAOST = 19, pPazAB = 20, ΔpazA = 25, and WT = 22). **(B)** Photos of *A. thaliana* at 14 d post inoculation display the differences in root phenotypes linked to bacterial inoculation by different *P. azotoformans* strains.

Supplementary Figure 2-21. Deuterium labeling supports PazB-mediated cyclization

Supplementary Table 2-12. Strains, plasmids, oligonucleotides, and DNA sequences used in this study

Supplementary Table 2-13. Summary of PazB variants

2.6 Conclusion

When studying plant root systems there are numerous barriers to consider. A range of physical, physiological, and methodological limitations hindering our understanding have been identified, as well as progress towards addressing some of these challenges.

The subsequent highlighted co-authored publications represent the development and application of such methods: Metabarcoding and data analytics programs to inform the basis of rhizosphere microbiome heritability in *Sorghum bicolor*, constitutive promoters to fine tune gene expression in synthetic biology that function in root systems, and utilizing root system imaging methods and software to inform the effect of a bacterial metabolite on plant-bacterial associations.

The improvements which enabled these studies have already yielded novel insights into root systems biology. Many challenges remain, but the recent consolidation of efforts across the discipline to standardize definitions and methods, along with technological advancements, are helping to overcome these barriers and allow for the collection of new kinds of data. These developments and their continuation will enable researchers to enhance the scope, reliability, and comparability of root system studies, ultimately contributing to a deeper understanding of plant biology and the many ways it intersects with the environment. It is our hope that these findings will lead to improvements in plant performance under new and changing agricultural practices better suited for the world we live in today.

Chapter 3. A rhamnosyltransferase mutant exhibits altered pectin monosaccharide composition in the cell wall of Lotus japonicus roots

3.1 Abstract

Pectin is an abundant component of cell walls demonstrated to be important in cell-to-cell adhesion, plant development, and a variety of signaling pathways. Understanding the underlying mechanisms of pectin biosynthesis and function has so far remained limited due to the wide diversity and redundancy of enzymes involved in the process. The component rhamnogalacturonan I (RG-I) is biosynthesized in part by rhamnosyltransferases (RRTs) in the GT106 family. While it is known that RRTs are expressed in a range of tissues, few studies have demonstrated their function outside of RG-I biosynthesis in seed coat mucilage. In this study we show *Lotus japonicus* RRT1 contributes to the addition of rhamnose monosaccharide residues to RG-I in root tissues. Mutants contained ~19% less rhamnose in their roots and pectin from aboveground tissues had less galactose and more xylose. RG-I in root tissue from *Ljrrt1* also had a larger molecular weight and altered structure compared to wild type (WT) Gifu plants, but this was not due to transcriptional differences in other GTs responsible for RG-I biosynthesis. Mutants exhibited altered root morphology, impacted stem and root growth, and impairment of nodule formation when inoculated with *Mesorhizobium loti*. We discuss potential implications of these findings and areas of future study.

3.2 Introduction

Pectin is an abundant component of cell walls demonstrated to be important in cell to cell adhesion, plant development, and a variety of signaling pathways.(492) Understanding the underlying mechanisms of pectin biosynthesis and function has so far remained limited due to the wide diversity and redundancy of enzymes involved in the process.(493–495) Advancements could provide useful information for the breeding and engineering of plants to develop root systems more amenable to sustainable agricultural practices, carbon sequestration strategies, biosynthesis of secondary products, and biofuels.(495)

Pectin contains three main components: homogalacturonan (HG), substituted galacturonan rhamnogalacturonan (RG-II), and rhamnogalacturonan I (RG-I).(494) Each of these require a series of glycosyltransferases (GTs) for their biosynthesis, but those required for RG-I biosynthesis are relatively recent additions to our understanding.(494,496,497) RG-I consists of a repeating sequence of [2)-α-L-Rha(1-4)-α-D-GalA(1-] disaccharides which are branched at the rhamnose residues with arabinan, galactan, or arabinogalactan.(498,499) The addition of rhamnose residues to the non-reducing end of RG-I is done by rhamnosyltransferases (RRTs), members of the GT106 family conserved across land plants.(493,496,500) The high level of RRT functional redundancy has made their study difficult, with many plants having a large number of *RRT* genes in their genome.(493) While it is known that RRTs are expressed in a range of tissues, few studies have demonstrated their function outside of seed coat mucilage.(493,496,500,501)

In this study we show *Lotus japonicus* RRT1 contributes to the addition of rhamnose monosaccharide residues to RG-I in root tissues. Mutants contained ~19% less

rhamnose in their roots and pectin from aboveground tissues had less galactose and more xylose. RG-I in root tissue from *Ljrrt1* also had a larger molecular weight and altered structure compared to wild type (WT) Gifu plants, but this was not due to transcriptional differences in other GTs responsible for RG-I biosynthesis. Mutants exhibited altered root morphology, impacted stem and root growth, and impairment of nodule formation when inoculated with *Mesorhizobium loti*.

3.3 Materials & Methods

3.3.1 Identification of candidate genes in LotusBase

LotusBase was used to BLAST the peptide sequence for *At*RRT1.(502,503) The four homologs (LotjaGi1g1v0209000, LotjaGi2g1v0270900, LotjaGi4g1v0257400, LotjaGi2g1v0150300) were checked for transcriptional across tissues and stimuli using the ExpressionAtlas. LjRRT1 (LotjaGi1g1v0209000) was selected for further characterization.

3.3.2 Germination and Growing Conditions of *Lotus japonicus*

Seeds were scarified in ≥98% concentrated sulfuric acid for 25 (Gifu) or 12 (*Ljrrt1)* minutes at 28°C and 600 rpm before 5 washes with deionized (DI) water. Seeds were surface sterilized in 10% bleach v/v for 2 minutes while shaking by hand before another 5 DI water washes. A final volume of DI water was added and seeds were left to imbibe while rotating at room temperature for 2-4 hours. Afterwards, they were placed on 1/2 Murashige and Skoog (MS), 1% w/v plant tissue culture agar (PhytoTech Labs – A111) plates with moistened filter paper to maintain humidity. The plates were sealed with parafilm and placed in the growth chamber (Percival Scientific: Model AR-100L3) set to 16:8hr light:dark regime (Hi Point Z4 Control LED Sunlight), 22°C, and 60% relative humidity for 7 days. Then seedlings were transplanted to 4.75-inch cone-tainers (Ray Leach Stuewe and Sons) filled with 0.3 cm^3 rockwool at the base, then 27.5 mL medium grain sand (Cemex Lapis Lustre Specialty Sands NO#60), and topped with 10 mL fine play sand (SAKRETE).

For uninoculated experiments, plants were watered with a 1/2 MS nutrient solution, covered with cling wrap and a humidifier dome, then placed in the growth chamber. They remained covered for 1 week and were then watered with 1/2 MS once a week for the duration of the experiment.

For inoculated experiments, plants were watered with a 1/2 MS without nitrogen solution, covered with cling wrap and a humidifier dome, then placed in the growth chamber. They remained covered for 1 week before flood inoculation with 3 mL of 0.1 OD_{600} *M. loti* solution per plant and were then watered once a week with 1/2 MS without nitrogen for the duration of the experiment.

3.3.3 Identification of homozygous *Ljrrt1* mutants

L. japonicus Gifu Wild Type (WT) and R3 generation LORE1 transposon mutagenized seeds were obtained from LotusBase. The LORE1 lines 30008895 (*Ljrrt1-1)* and 30109270 (*Ljrrt1-2*) were germinated and grown for 7-14 days as described.(504) Homozygous transposon insertions in RRT1 were confirmed using the Plant Phire

Direct (ThermoFisher #F-160S) and primers (Supplemental Table 3-1) as described in LotusBase. Those identified were kept for seed bulking in the previously described growing conditions.

3.3.4 Alcohol Insoluble Residue (AIR) preparation & TFA Hydrolysis

Harvested tissues were placed in 2 mL screw cap tubes with 95-100% ethanol and heated to 100°C for 30 minutes to halt enzymatic activity. The Powerlyzer 24 Homogenizer (Qiagen) was used at the recommended settings for corresponding tissues to lyse tissue. Beads were removed and samples were centrifuged at 14,000rpm for 7 minutes. The supernatant was discarded and then the pellet was resuspended in 95-100% ethanol before centrifuging again. Ethanol was repeated with 70% ethanol until the supernatant was colorless. The pellet was washed in acetone before being left to dry at 50°C overnight.

10 mg of AIR preparation for each sample were weighed then resuspended in 1 mL 2M trifluoroacetic acid (TFA) (Sigma Aldritch) and heated at 120°C for 1 hour with occasional vortexing. TFA was removed via acid cold-trap speed vacuum overnight. The pellet was then resuspended in 1 mL milliQ water and shaken at 1000 rpm, 30°C for 1 hour before being filtered through 0.45 um nitrocellulose (VWR). Filtrate was diluted 1:10 with water and adjusted to a pH of 4.0 – 9.0 for further analysis using High Performance Anion Exchange Chromatography.

3.3.5 Pectin Enrichment of Hydrolyzed AIR Preparations and Size Exclusion Chromatography

RG-I was isolated essentially as described by (505) with some modifications. Briefly, AIR preparations were weighed out and placed into 2 mL screw cap tubes with three metal beads and 0.5 mL of 50 mM ammonium oxalate (pH 5.0) before using the Powerlyzer 24 Homogenizer (Qiagen) recommended settings. Beads were removed and 1 uL (20U) endo-polygalacturonase (Megazyme, product code E-PGALUSP), 1 uL (3U) Novoshape Pure (Novozymes, Redding University), and 600 uL of ammonium oxalate was added. Samples were left to digest at 37°C, 600 rpm overnight (~12 hours). After being centrifuged at max speed for 5 minutes, supernatant was filtered through 0.45 um nitrocellulose (VWR) (14,000 rcf for 2 minutes). 10 kDa Molecular Weight Cutoff concentrator columns (Amicon) were used to wash away oligosaccharides and the digestion buffer according to the manufacturer instructions. Samples were eluted from the spin concentrators in water and analyzed using HPAEC and Size Exclusion Chromatography.

3.3.6 High Performance Anion Exclusion and Size Exclusion Chromatography

Non-cellulosic monosaccharide composition was detected and quantified using High Performance Anion Exchange Chromatography with Pulsed-Amperometric Detection (HPAEC-PAD) using a Thermo Scientific Dionex ICS-6000 system as described by (506). The column was equilibrated with 5mM NaOH for 5 min. Neutral sugars were separated over a 5 mM sodium hydroxide 0.4 ml/min gradient over 23 minutes before separating the uronic acids using 450 mM sodium hydroxide at 0.4 ml/min over 23 - 41 minutes using a Dionex CarboPac PA20 column (3 x 30 mm, 060144). Amounts were quantified using a range of monosaccharide standards (2.5-200 µM).

RG-I and RG-II were separated and analyzed by size-exclusion chromatography in 50 mM ammonium formate (pH5.0) on a Superdex 200 10/300GL column (Cytiva) as described (507). Elution of polysaccharides at a flow rate of 0.5 ml/min from the column was monitored with a Shodex RI-101 refractive index detector (Shodex, http://www.shodex.com). Estimates of the MW of RG-I were made with reference to the retention times of Dextran MW standards (Sigma-Aldrich). The relative mass percentage of each monosaccharide in the RG-I fractions was determined by first calculating the mass ratio of each monosaccharide by dividing the product of the mol% and the molar mass of each monosaccharide by the sum of the products of the molar mass and mol% of each monosaccharide. We then multiplied the monosaccharide mass ratios by the estimated molecular weight of the RG-I fraction.

3.3.7 RNA extraction and quantitative real-time PCR of target genes

To quantify expression of target genes, 100 mg of 28 dpt roots from nutrient sufficient conditions were flash frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) and corresponding DNase. Complementary DNA synthesis was conducted using the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) from 500 ng of total RNA, and quantitative polymerase chain reaction (qPCR) was conducted from cDNA diluted 1:5 using the PowerUp SYBR Green Master Mix (ThermoFisher Scientific). A 200 nM primer concentration and the following protocol were used for qPCR for all targets: 2 min at 50 °C and 2 min at 95 °C, followed by 39 repeats of 15 s at 95 °C, 15 s at 60 °C and 1 min at 72 °C, and ending with 5 s at 95 °C. A melting curve (55–95 °C; at increments of 0.5 °C) was generated to verify the specificity of primer amplification. four 28 dpt biological replicates and three technical replicates of all targets (*LjARAD1* and *LjGALS1*) were quantified for gene expression levels relative to the reference gene, a *L. japonicus* polyubiquitin (LotjaGi5g1v0317900), using the ΔΔCT method. All primer sequences used for qPCR can be found in Supplementary Table 3-1. Raw ΔΔCT values used for statistical analysis can be found in Supplementary Table 3-2.

3.3.8 Sample preparation and data collection for physiological phenotypes

For uninoculated, nutrient sufficient experiments, following transplanting samples were harvested at desired time points by hand, roots were cleaned with DI water, patted dry with a paper towel, then weighed and images were taken for later measurement in ImageJ. A subset of samples were split into above- and belowground portions and then left to dry at 50°C for 48 hours before collecting dry weight.

For *M. loti* inoculated experiments, plants were harvested and observed with an Olympus SZX16 stereoscope fitted with an Olympus SZX2-ILLT base for transmitted/oblique illumination to assess (im)mature nodule formation. Plants remained in water to prevent desiccation, then patted dry before following data collection. Sample preparation and data collection were otherwise the same.

3.3.9 Preparation of *Mesorhizobium loti* for Plant Inoculation

Mesorhizobium loti R7A was grown for 48hrs at 28°C, 180 rpm in Tryptone-Yeast broth. The cultures were then transferred to 15 or 50 mL labeled falcon tubes and pelleted using a swing

arm centrifuge at 4000 rcf for 10-15 min. The media was decanted and bacteria were resuspended in DI H2O before two rounds of washing. The bacteria were resuspended a final time in DI H20 and adjusted to 0.1 OD_{600}.

3.3.10 Data Analysis

All quantitative data analysis and visualization occurred in Jupyter notebook with custom Python scripts using the SciPy, NumPy, Pandas, Matplotlib, and Seaborn libraries. The Analysis of Variance (ANOVA) statistical test was used for all experiment data.

3.4 Results

3.4.1 *Lotus japonicus* contains four RRT homologs with varied expression patterns

Figure 3-1. LotusBase ExpressionAtlas profiles for the four *L. japonicus AtRRT* homologs

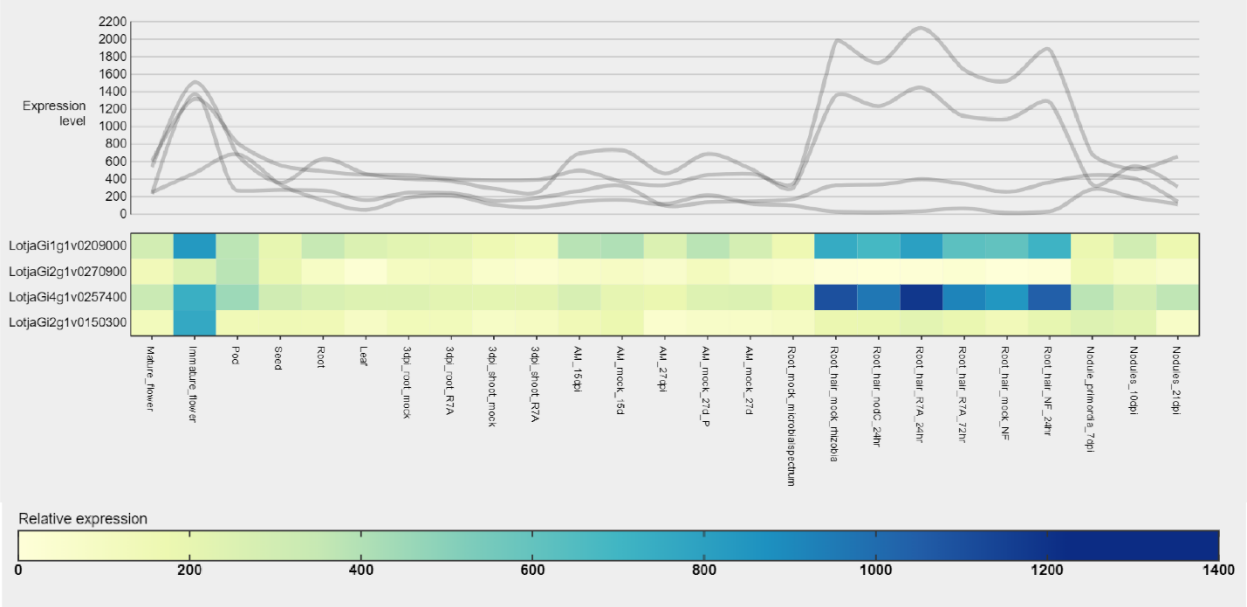


Figure 3-1. LotusBase Expression Atlas line and heat map displaying the measured expression levels of the four *AtRRT* homologs identified via peptide sequence similarity. Selection for further study was based on both overall expression levels across tissues and conditions as well as levels in root tissue specifically. Data is compiled from multiple experiments.

Utilizing the BLAST and ExpressionAtlas features of LotusBase, we determined *L. japonicus* contains four genes homologous to *Arabidopsis thaliana RRTs* which had differing transcriptional expression patterns across tissues and stimuli.(496,502–504) Transposon knock out mutants of *LjRRT1* (LotjaGi1g1v0209000.1;LORE1 ID:30008895) were used for further investigation due to having the highest transcriptional activity in roots and dynamic expression (Figure 3-1).

3.4.2 *Lotus japonicus rrt1* mutants exhibit a 15% reduction of rhamnose in root tissue pectin

Figure 3-2. Differences in root tissue cell wall rhamnose content of *Ljrrt1* mutants

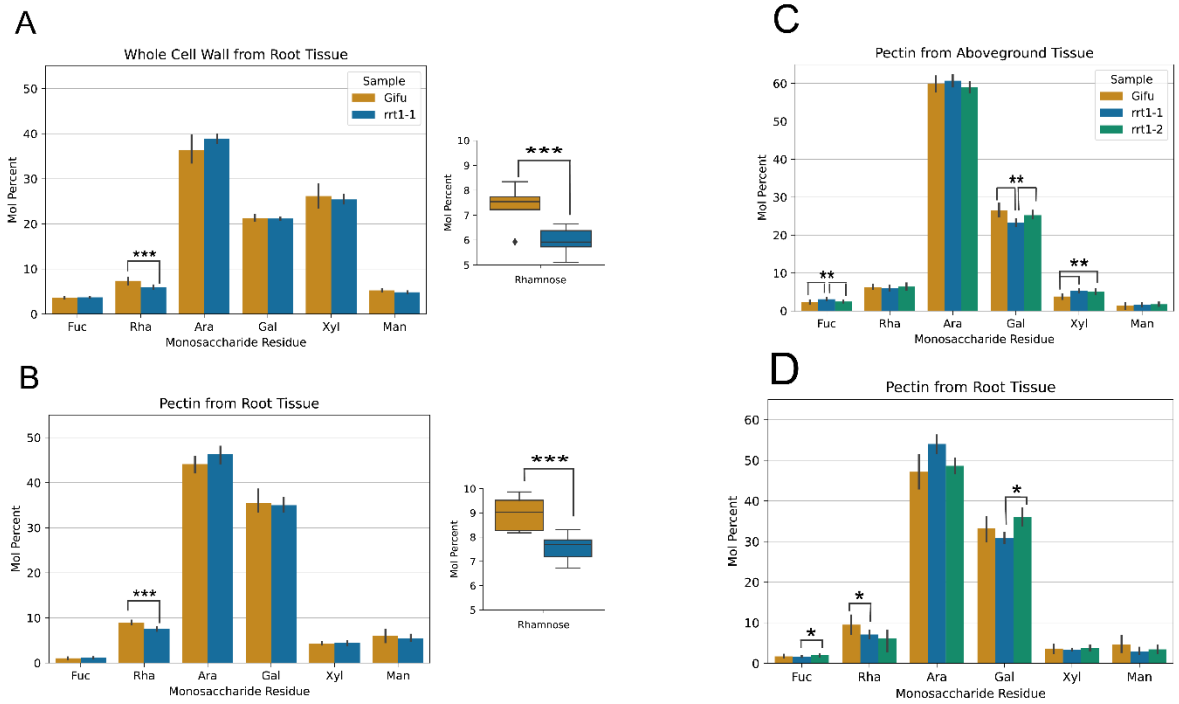


Figure 3-2. Neutral monosaccharide composition from alcohol-insoluble residue (AIR) preparation, as measured by HPAEC and expressed as Mol percent of: **(A)** the entire cell wall portion and **(B)** pectin enriched fraction of 28 dpt root tissue from Gifu WT and *Ljrrt1-1* (n = 5 and 6, respectively); **(C)** pectin enriched fraction of 28 dpt aboveground and **(D)** root tissue from Gifu WT, *Ljrrt1-1*, and *Ljrrt1-2* (n = 9, 9, and 6 respectively). P-values for rhamnose content compared to WT in **D** are 0.08 (*Ljrrt1-1*) and 0.15 (*Ljrrt1-2*). Fuc = fucose, Rha = rhamnose, Ara = arabinose, Gal = galactose, Xyl = xylose, Man = mannose. P-values are indicated as *** ≤ 0.02, ** ≤ 0.05, * ≤ 0.1 and determined by Analysis of Variance (ANOVA).

L. japonicus rrt1-1 plants had approximately 19% less content of rhamnose residues in their root cell walls than Gifu WT (Figure 3-2A); mainly due to the pectin fraction, where an approximately 16% reduction was seen (Figure 3-2B). This reduction resembles *rrt* mutants in *Marchantia polymorpha*, implying similar functional redundancy among other RRT homologs in the genome.(493) *Ljrrt1-1* pectin also tended to have slightly higher arabinose content, but the degree of significance varied between experiments (Figure 3-2B,D). Interestingly, pectin from aboveground tissues in *Ljrrt1-1 and Ljrrt1-2* mutants had no reduction in rhamnose but did show an increase in xylose content (+40% and +32% respectively) and *Ljrrt1-1* had a reduction in galactose (13%) (Figure 3-2C).

3.3.3 RG-I in root tissue from *rrt1* mutants has an increased molecular weight and altered structure

Figure 3-3. Size exclusion elution time, calculated molecular weight, and monosaccharide content indicates RG-I from *Ljrrt1* root tissue has elongated galactan and arabinan side chains

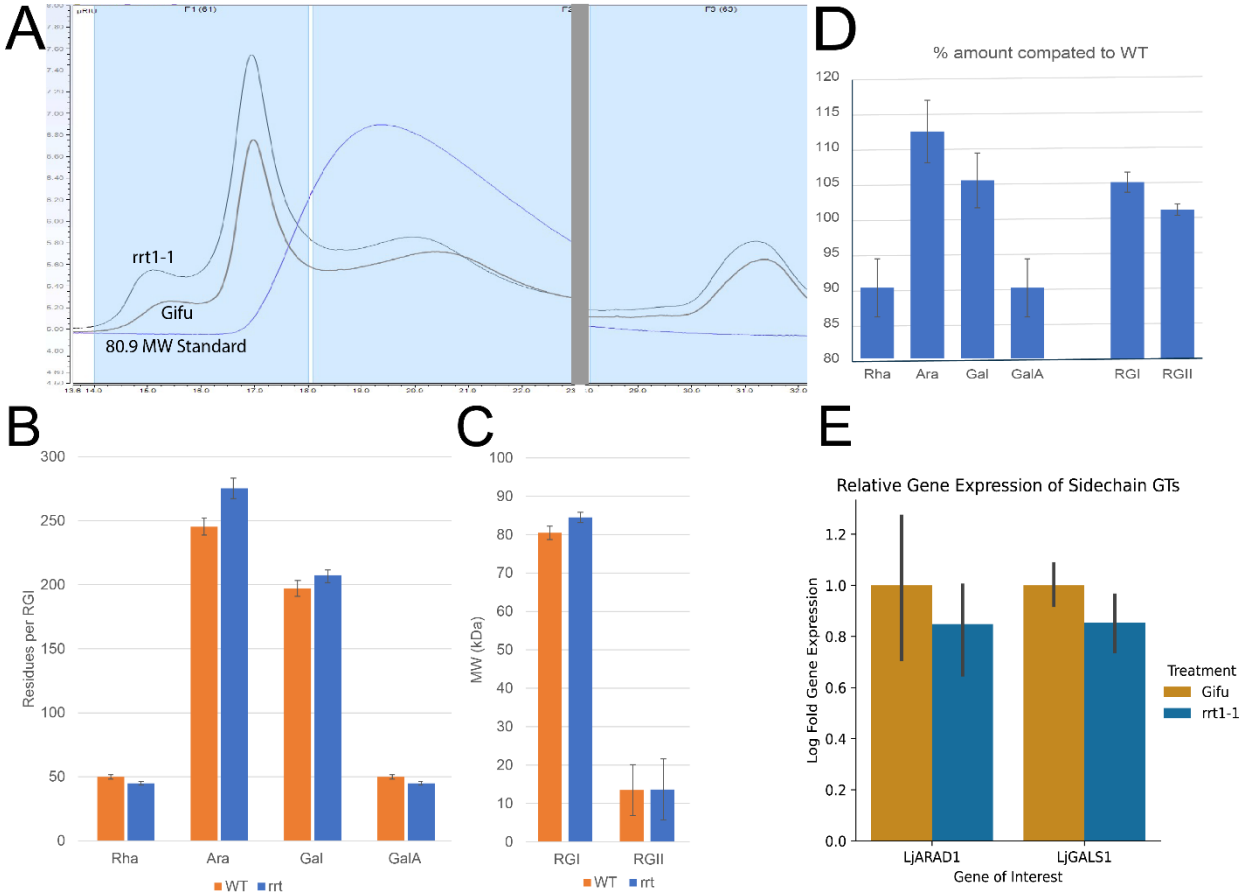


Figure 3-3. The pectin fraction of *Ljrrt1-1* root tissue was analyzed via size exclusion chromatography **A)** Chromatograms displaying Gifu WT and *Ljrrt1*-1 RG-I elution profiles alongside an 80.9 MW standard. A break was added to highlight the observed peaks **B)** Number of residues per RG-I and **C)** Molecular weight (kDa) of RG-I and RG-II **D)** Relative percentages of monosaccharide residues, RG-I, and RG-II (n = 8) **E)** Expression of *LjARAD1* and *LjGALS1* in root tissue 28 dpt measured by qRT-PCR relative to *LjUBQ* (LotjaGi5g1v0317900); mean ± SEM (n = 4)

RG-I in root tissue from *Ljrrt1-1* mutants also had an altered molecular mass and structure. Mutant-sourced RG-I had a longer size exclusion elution time than RG-I from Gifu WT (Figure 3-3A). Calculations using the determined molecular weight and monosaccharide mol percentages indicated mutant RG-I was 5% larger due to a greater number of galactose and arabinose residues, which would constitute longer galactan and arabinan sidechains, while no such differences were observed for RG-II (Figure 3-3B-D). These side chains are formed by GTs in the GALACTAN SYNTHASE (GALS) and ARABINAN DEFICIENT (ARAD) families.(508,509) The respective *L. japonicus* homologs, determined by highest peptide sequence similarity, were selected to determine if changes in transcriptional regulation could explain the elongated side chains. However, qPCR results indicated this was not attributed to a change in the

expression of *LjGALS1*(LotjaGi6g1v0246500) or *LjARAD1*(LotjaGi4g1v0082700) (Figure 3-3E).

3.4.4 *Ljrrt1* mutants have stunted growth and altered root system architecture and morphology in nutrient sufficient conditions

Figure 3-4. Physiological differences between *L. japonicus* Gifu and *rrt1* plants

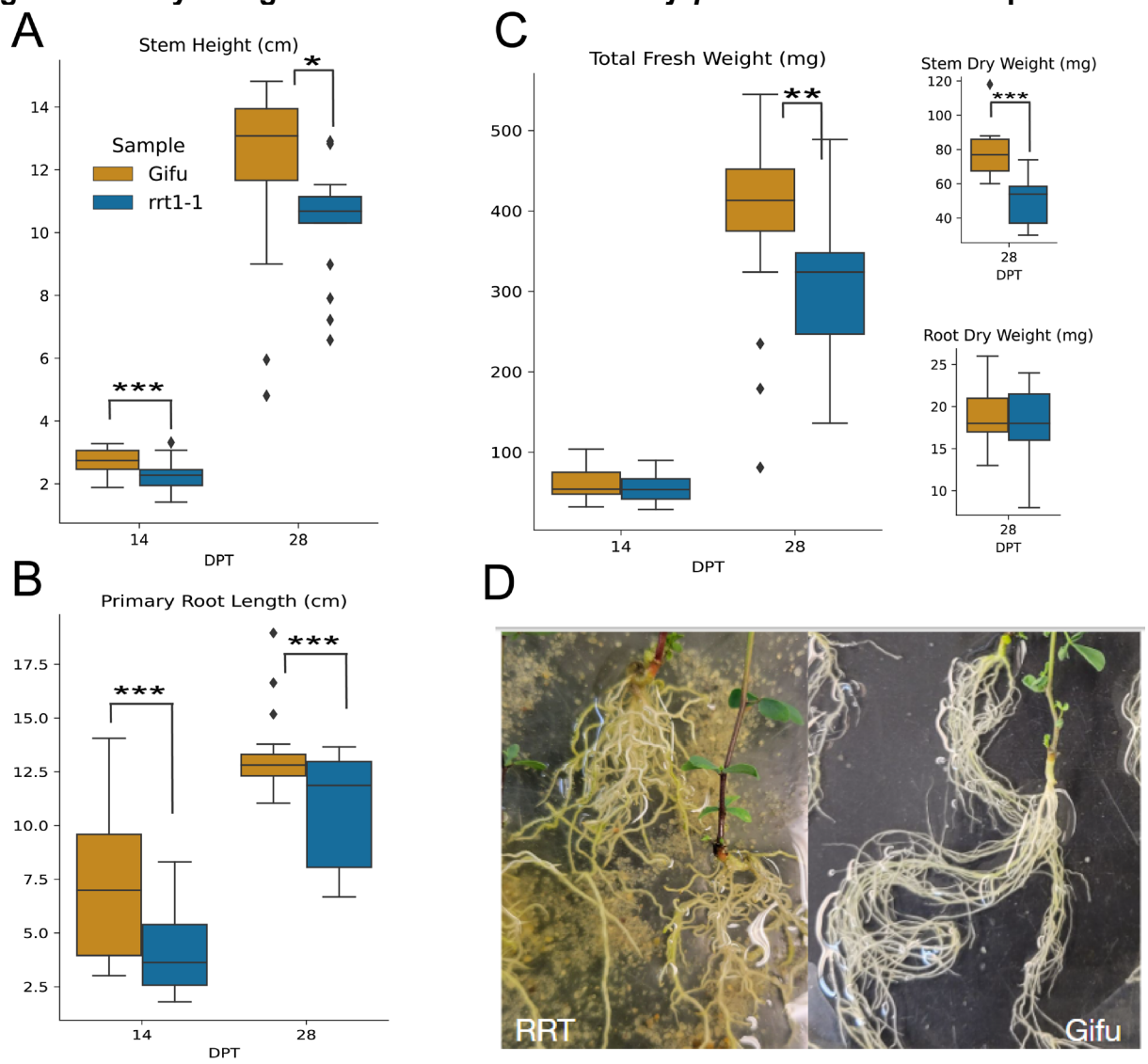


Figure 3-4. Gifu WT and *Ljrrt1-1* plants were grown in sand and watered with 1/2 Murashige and Skoog nutrient media. Samples were collected at 14- and 28-dpt and measured for **A)** stem height (cm) **B)** primary root length (cm) (n =16) **C)** fresh and dry weight (mg) (n = 7) **D)** Images depicting *Ljrrt1-1* and Gifu WT root morphology. P-values are indicated as *** ≤ 0.02, ** ≤ 0.05, * ≤ 0.1 and determined by ANOVA.

L. japonicus rrt1-1 plants grown in nutrient sufficient conditions were shorter in stem height (17% and 14%) and had less primary root growth (43% and 18%) at 14- and 28-days post-transplant, respectively (dpt) (Figure 3-4A,B). However, there was no difference in total fresh weight until 28 dpt, where stems appeared to be responsible for the observed 20% weight reduction (Figure 3-4C). Lj*rrt1-1* mutants also had thicker

roots and altered root system architecture (Figure 4-D). During the germination protocol, the seeds scarified with acid roughly twice as fast as WT seeds. The standard scarification protocol led to very low germination, possibly because of less protection from the strong acid by pectin in the cell walls.

3.4.5 *Ljrrt1* mutants grown in nitrogen-limited conditions show exacerbated growth impacts and have slight impairments in nodule formation with *M. loti*

Figure 3-5. Physiological data for Gifu and *Ljrrt1* mutants in nitrogen-limited conditions: smaller, but not lighter, plants that form less mature and immature nodules at different time points

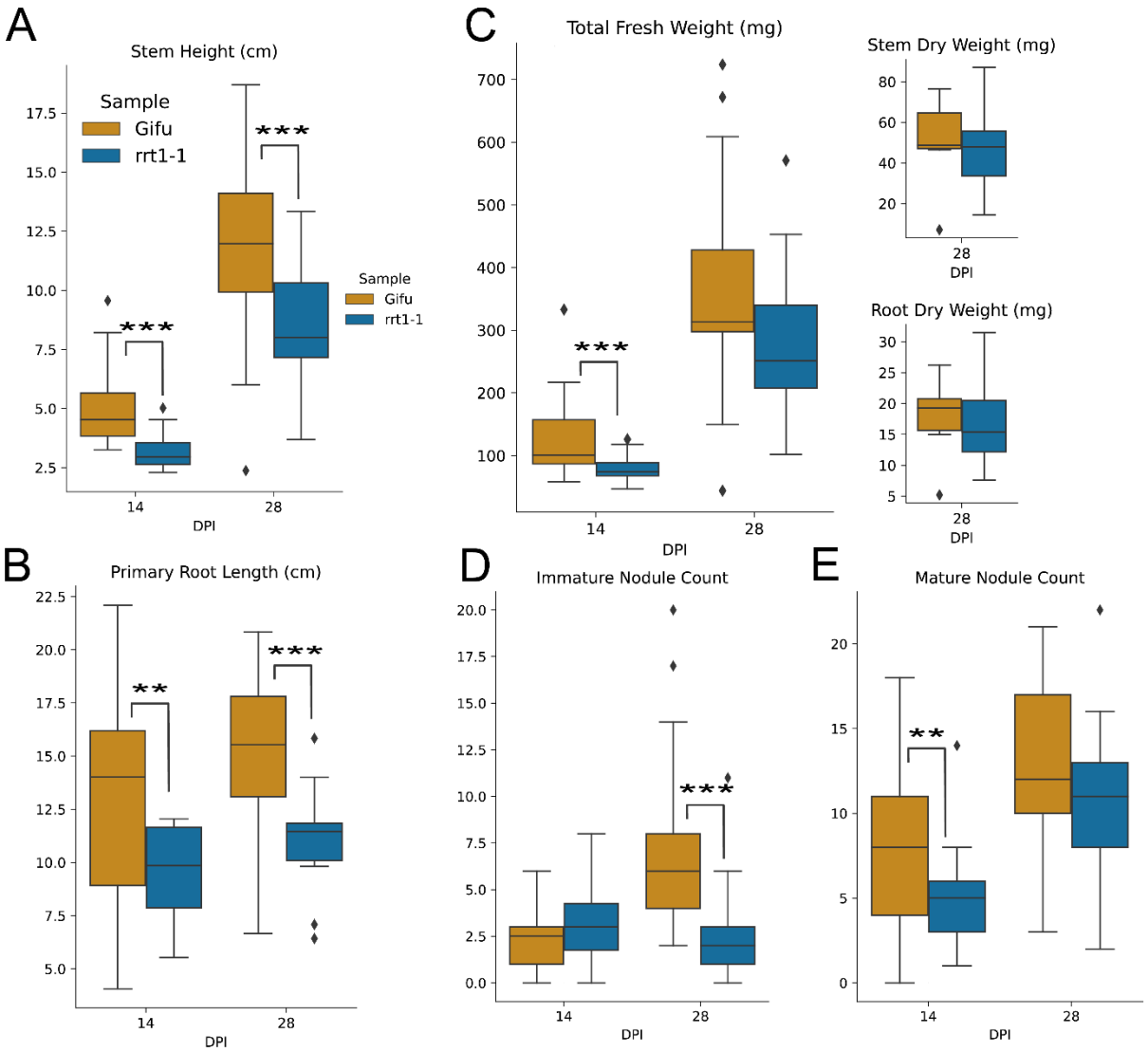


Figure 3-5. Gifu WT and *Ljrrt1-1* plants were grown in sand, inoculated with *M. loti* and watered with 1/2 Murashige and Skoog nutrient media without nitrogen. Samples were collected at 14- and 28-dpt and measured for **A)** stem height (cm) **B)** primary root length (cm) (n =16) **C)** fresh and dry weight (mg) (n = 7) **D)** immature nodule and **E)** mature nodule count (n =16). P-values are indicated as *** ≤ 0.02, ** ≤ 0.05, * ≤ 0.1 and determined by ANOVA.

RG-I may influence the formation of bacterial associations in root systems through its involvement with mucilage production and cell wall modification.(496,510,511) *L. japonicus rrt1-1* plants grown in nitrogen-depleted conditions and inoculated with *M. loti* demonstrated impacted growth and nodule formation. The 25% reduction in primary root length for the duration of the experiment was similar to that in nutrient sufficient, uninoculated conditions, however the reduction in stem height was exacerbated (36% and 25% at 14 and 28dpi) (Figure 3-5A,B). The fresh and dry weights of *Ljrrt1-1* mutants and Gifu WT plants showed little difference outside of 14 dpi, despite the increased impact on growth (Figure 3-5C). When documenting the progression of rhizobial symbiosis, *Ljrrt1-1* mutants had formed roughly 30% less mature nodules at 14 dpi and 68% less immature nodules at 28 dpi (Figure 3-5D,E).

3.5 Discussion

L. japonicus was determined to possess four *At*RRT homologs based on peptide sequence similarity. Root tissue in *Ljrrt1* mutants exhibited an approximately 19% and 16% reduction of rhamnose monosaccharide content in the whole and pectin-enriched cell wall portions respectively, similar to results from *Marchantia polymorpha rrt1* knockout mutants (Figure 3-2).(493) This demonstrates the RRT clade's involvement in RG-I biosynthesis outside of seed coat mucilage, as well as reconfirms the functional redundancy of RRT seen in other studies.(493,496) Whether this redundancy is achieved via transcriptional up-regulation or biochemical activity of other RRT homologs, or perhaps other GTs, is uncertain. GT family 106 is a large family with e.g. 34 members in Arabidopsis. While some members of 106 are known not to have RRT activity, it is currently unclear if RRT activity is restricted to the small RRT clade or is also present in other GT106 subclades. High sequence similarity between *LjRRT1-4* prevented accurate assessment of transcript levels via qPCR. Interestingly, pectin from aboveground tissues of *Ljrrt1* mutants had no differences in rhamnose but showed a reduction in galactose and an increase in xylose content (Figure 3-2C). Though not measured, this suggests that RG-I in these tissues did not undergo the same changes seen in roots and the molecular mass of RG-I may actually be lower compared to the WT, in contrast with the roots where we determined a 5% larger molecular mass of RG-I. How this manifested is unclear, but it may indicate complexity in the balancing of saccharide pools across tissues for cell wall biosynthesis. The increased xylose content could indicate an increased biosynthesis of xylogalacturonan in the shoots, as opposed to increased RG-I molecular mass in the roots.

RG-I in *Ljrrt1* mutants was calculated to have an altered structure compared to Gifu WT plants, with longer galactan and arabinan side chains and higher overall molecular mass. There were no observed differences in the expression of *LjGALS1* and *LjARAD1*, ruling out transcriptional regulation as a possible cause for these longer side chains. Other studies have shown that alteration to galactan side chains can change interactions with cellulose by displacing xylan.(512) This finding may help shed light on how alterations to monosaccharide ratios can lead to larger changes in cell wall structure *in vivo*, with potential applications for product biosynthesis and biofuels.

The altered cell wall composition of *Ljrrt1* mutants translated into several physiological phenotypes. They exhibited shorter stem heights, possibly due to increased xylose and reduced galactose residues in the pectin of aboveground tissues (Figure 3-2C, Figure 3-4A). Mutants also had a reduction in primary root length and altered root system architecture and morphology (Figure 3-4D). The same was seen in nitrogen limited, *M. loti* inoculated conditions, with a greater reduction in stem height (Figure 3-5A). There were also reductions in mature and immature nodule formation at 14 and 28 dpi, respectively (Figure 3-5).

The shorter stem height may be explained by differences in pectin composition as well as less overall surface area in the root system impacting nutrient foraging and uptake. However, further experimentation is required to determine if this is the case. Reduction in rhamnose content of the cell wall leading to changes in root morphology has been observed in non-functional RHAMNOSE BIOSYNTHESIS 1 (RHM1) *Arabidopsis thaliana* mutants which are impacted upstream of RRT function.(510) Additionally, other varieties of rhamnosyltransferases have been shown to add rhamnose residues to secondary metabolites, particularly flavonoids, demonstrated to have a variety of protective effects from (a)biotic factors and influence in developmental processes.(513). It is possible that a lower demand for rhamnose for pectin biosynthesis would result in increased rhamnosylation of secondary metabolites. Whether the changes observed in our study are due to altered cell wall structure, alterations to processes via changes to rhamnosylated secondary metabolite pools, or a combination of these factors is currently undetermined. More investigations concerning this question could be useful in understanding the nuanced dynamics in the interplay between cell wall modification and cellular signaling.

There are a multitude of factors that could contribute to the observed impacts on nodule formation: less overall root surface area and nutrient uptake, possible root hair deformation, altered cell wall structure inhibiting lateral root and infection thread development, and altered flavonoid production all are hypotheses to be explored. Considering the role of flavonoids, root hairs, and infection threads in establishing rhizobial symbiosis, further study may clarify connections between cell wall modification and signaling in cell development as it pertains to endosymbiosis.(514) However, the redundancy of RRT function remains a hurdle in unraveling such questions through limiting the development of rhamnose-deficient mutant phenotypes.

3.6 Conclusion

Our study has demonstrated root tissue specific RRT function in RG-I biosynthesis, with an approximately 19% and 16% reduction of rhamnose content in the whole cell wall and pectin from *L. japonicus rrt1* transposon knockout mutants. Despite this reduction, RG-I from mutant root tissue was calculated to have longer galactan and arabinan side chains and higher molecular weight, which could not be explained by changes in expression of related GTs. Pectin from aboveground tissue of *Ljrrt1* mutants had a reduction in galactose and increase in xylose, with no changes to rhamnose content. Additionally, *Ljrrt1* mutants exhibited physiological defects in root system architecture, growth, and morphology as well as minor effects on their ability to form nitrogen-fixing

nodules with *M. loti*. Further study can reveal fundamental insights into cell wall biosynthesis and the varied roles of RG-I in plant development and response to stimuli.

3.7 Acknowledgements

This work was not possible without the assistance of many people. Tomo Yoshino, Thalissa Malagoli Franzon, and Victoria Vera assisted in conducting experiments, collecting data, and data analysis. Yi-chun Chen provided technical expertise in HPAEC and size exclusion chromatography methods. Dr. Henrik Scheller assisted in experiment design, data analysis, and manuscript editing. This work was supported by the InRoot Project, which is funded by the NovoNordisk Foundation, & conducted at the Joint BioEnergy Institute which is supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Lorenzo Washington was further supported by the National Science Foundation under Grant No. DGE 2146752 & the Berkeley Fellowship for Graduate Study.

3.8 Supplemental Information

Located in Appendices

Supplemental Table 3-1. Primers used in study

Supplemental Table 3-2. Raw log-fold ddCt values for qPCR

Chapter 4. Small Extra-Large GTPase-like proteins influence rhizobial symbiosis in Lotus Japonicus

4.1 Abstract

Plants possess a unique class of heterotrimeric Gα subunits called extra-large GTPases (XLGs) which contribute to numerous developmental and stress responses. XLGs have an uncharacterized N-terminal domain, a Gα-like C-terminal domain, and overlapping and distinct functions compared to conventional Gα subunits. In this study, we identified homologs of XLG3 in *Lotus japonicus* responsive to rhizobial and mycorrhizal symbiosis. However, these proteins were approximately one-third the size of conventional XLGs and only aligned to the N-terminal domain, containing a putative NLS and the cysteine-rich domain of unknown function. Multiple sequence alignment and phylogenetic analysis determined SXLGs did not share domains with other mono- or heterotrimeric G-protein classes and exhibited a pattern of duplication and neofunctionalization typical of genes involved in symbiotic signaling pathways. Transient expression of *LjSXLG*s in tobacco demonstrated their potential for localization to the plasma membrane, nucleus, and nucleolus. Analysis of *L. japonicus sxlg2* mutants revealed transient impairment of immature nodule formation in a destructive experimental setup and inhibition of infection events in a nutrient-limited non-destructive experimental setup, with no observed difference in nodule maturation rate. Additionally, *sxlg2* mutants showed a potential impairment of the root growth response in N-limited conditions. We discuss the potential utility of SXLGs in better understanding XLGs evolution and function.

4.2 Introduction

Plants possess the heterotrimeric guanosine triphosphate (GTP)-binding protein as a means of relaying extracellular signals into their downstream pathways.(515) Similar to animal and fungal systems, this heterotrimeric protein is composed of Gα, Gβ, and G**γ** subunits which disassociate upon activation to relay signal transduction; however plants possess distinct elements such as alternative regulation strategies, unique G**γ** subunit types, and an additional class of Gα subunits in Extra-Large Gα-proteins (XLGs).(515,516) XLGs are composed of an uncharacterized N-terminal domain, possessing a nuclear localization signal (NLS) and cysteine-rich domain of unknown function (DUF), and a C-terminal Gα-like domain.(516,517) They have been demonstrated to provide the diversity that was previously thought to be lacking in plant Gα subunits, with distinct and overlapping roles and functions compared to conventional Gα subunits.(515,517,518)

G-proteins have been found to be involved in a wide variety of plant signaling pathways and utilize the diversity present amongst subunit classes to effectively respond across such a range of signals.(517–521) Interestingly, despite operating on a timescale of seconds to minutes, plant G-proteins have been primarily associated with slow-scale developmental responses, such as organ development and cell elongation.(522,523) However, they are also implicated to participate in responses to biotic interactions, such as in immunity and symbiotic association with nodule-inducing rhizobia, processes which involve rapid intracellular responses to external stimuli.(517,519,520,524) Additionally, XLGs were shown to be involved in numerous pathways that incorporate

environmental stresses, such as nutrient deficiency and infection, with developmental outcomes in above and below-ground tissues.(517,518,525)

This places XLGs as a potentially fruitful site of investigation regarding symbiotic associations in plant root systems, as these require the incorporation of numerous (a)biotic signals to determine large-scale developmental changes.(526,527) In this study, we used LotusBase to identify homologs of XLG3 in *Lotus japonicus* which exhibited expression patterns specific to stages of rhizobial and mycorrhizal symbiosis. However, these proteins — further referred to as small XLGs (SXLGs) — were approximately one-third the size of conventional XLGs and aligned to the N-terminal domain, possessing putative NLS and the cysteine-rich DUF. Multiple sequence alignment and phylogenetic analysis determined SXLGs did not share domains with similarly sized G-protein classes, such as Gβ and G**γ** subunits or small monomeric GTPases such as ROP, and exhibited a pattern of duplication and neofunctionalization typical of genes involved in symbiotic signaling pathways. Transient expression of SXLGs demonstrated their potential for localization to the plasma membrane, nucleus, and nucleolus. Analysis of *Lotus japonicus sxlg2* mutants revealed transient impairment of immature nodule formation in a destructive experimental setup and inhibition of infection events in a nutrient-limited non-destructive experimental setup, with no observed difference in nodule maturation rate. Additionally *sxlg2* mutants showed a potential impairment of the root growth response in N-limited conditions.

4.3 Materials & Methods

4.3.1 Identification of candidate genes in LotusBase, Multiple Sequence Alignment, and Phylogenetic Tree Generation

LotusBase was used to BLAST the peptide sequence for *Lj*XLG3.(502,503) The top ten homologs were checked for transcriptional expression responses to symbiotic conditions using the ExpressionAtlas (ExPat). Peptide sequences for *LjSXLG1*(LotjaGi1g1v0106400) and *LjSXLG2*(LotjaGi6g1v0043000) were aligned with Mega-X software: Muscle alignment and tree generation using UPGMA cluster methods.(528–531) The additional sequences were collected via pBLAST using the non-redundant protein database in NCBI.(532)

4.3.2 Germination and Growing Conditions of *Lotus japonicus*

Seeds were scarified in ≥98% concentrated sulfuric acid for 25 minutes at 28°C and 600 rpm before 5 washes with deionized (DI) water. Seeds were surface sterilized in 10% bleach v/v for 2 minutes while shaking by hand before another 5 DI water washes. A final volume of DI water was added and seeds were left to imbibe while rotating at room temperature for 2-4 hours. Afterwards, they were placed on 1/2 Murashige and Skoog (MS), 1% w/v plant tissue culture agar (PhytoTech Labs – A111) plates with moistened filter paper to maintain humidity. The plates were sealed with parafilm and placed in the growth chamber (Percival Scientific: AR-100L3) set to 16:8hr light:dark regime (Hi Point Z4 Control LED Sunlight), 22°C, and 60% relative humidity for 7 days before transplanting. All experiments using *Lotus japonicus* were done in these growing conditions. Wildtype Gifu and *Ljsxlg2* (LORE1 ID: 30063804; LotjaGi6g1v0043000) seeds were given the same treatment.

4.3.3 Identification of homozygous *sxlg2* mutants

L. japonicus Gifu Wild Type (WT) and R3 generation LORE1 transposon mutagenized seeds were obtained from LotusBase. The LORE1 line (30063804) seeds were germinated and grown for 7-14 days as described.(504) Homozygous transposon insertions in SXLG2 were confirmed using the Plant Phire Direct Kit (ThermoFisher #F-160S) and primers (Supplemental Table 4-1) as described in LotusBase. Those identified were kept for seed bulking in the previously described growing conditions.

4.3.4 Bacterial Strains and Culture

Mesorhizobium loti R7A and kanamycin-resistant *M. loti* pGingerRFP (red fluorescent protein) were grown at 28°C on tryptone-yeast (TY) media.(533) GV3101 *Agrobacterium tumefaciens* strains pCL2/pCM1/pNOS-mNeonGreen::*LjXLG1/2* were grown at 30°C on Luria Broth (LB) media with the 100 ug/mL rifamycin, 30 ug/mL gentamicin, and 50 ug/mL kanamycin.

4.3.5 RNA extraction and quantitative real-time PCR of target genes

To quantify expression of target genes, 100 mg of roots from the sand cone-tainer rhizobia-inoculated experiment were flash frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) and corresponding DNase. Complementary DNA synthesis was conducted using the SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific) from 500 ng of total RNA, and quantitative polymerase chain reaction (qPCR) was conducted from cDNA diluted 1:5 using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific). A 200 nM primer concentration and the following protocol were used for qPCR for all targets: 2 min at 50 °C and 2 min at 95 °C, followed by 39 repeats of 15 s at 95 °C, 15 s at 60 °C and 1 min at 72 °C, and ending with 5 s at 95 °C. A melting curve (55–95 °C; at increments of 0.5 °C) was generated to verify the specificity of primer amplification. Three (7dpi) and four (21dpi) biological replicates and three technical replicates of all targets (*LjSXLG1* and *LjSXLG2*) were quantified for gene expression levels relative to the reference gene, a *L. japonicus* polyubiquitin (LotjaGi5g1v0317900), using the ΔΔCT method. All primer sequences used for qPCR can be found in Supplementary Table 4-1. Raw ΔΔCT values used for statistical analysis can be found in Supplementary Table 4-2.

4.3.6 Plasmid Construction and Bacterial Transformation

All expression constructs were generated using PCONS plasmids possessing a GFP-dropout.(534) A monomeric NeonGreen fluorescent protein was fused to the C-terminus of the coding sequences for *LjSXLG1/2* via Gibson reaction (New England Biolabs HiFi Assembly). The fusions were then inserted via Golden Gate BSA1 restriction cloning (New England Biolabs) into two different strength PCONS (pCL2 and pCM1) and a similar plasmid using the nopaline synthase (NOS) promoter. Chemically competent XL1-Blue *E. coli* (QB3 MacroLab, USA) grown at 37°C were used in the construction steps before the final versions were electroporated into GV3101 *Agrobacterium tumefaciens*. α-DHA *E. coli* were used to generate *M. loti* expressing a pGinger RFP construct via conjugation. Briefly, respective bacteria were grown, pelleted, then resuspended and aliquoted as a mix on TY plates at 28°C; after 2-4 days the mixed

bacteria were streaked onto TY plates selective for the pGinger plasmid. Successful colonies were cultured, surveyed, and sequenced to determine conjugation success. Plasmids were routinely isolated using the Qiaprep Spin miniprep kit (Qiagen, USA), and all primers were purchased from Integrated DNA Technologies (IDT; Coralville, IA)

4.3.7 Transient Expression in *Nicotiana benthamiana* and Microscopy

Agroinfiltration protocol was adapted from Sparkes et al.(433) Transformed agrobacterium were grown in LB liquid media with 50 µg/mL kanamycin, 50 µg/mL rifampicin, and 30 µg/mL gentamicin to between optical density (OD) 0.6 and 1 before diluting to 0.05 OD_{600} (SXLG::mNeonGreen fusion) or 0.15 OD_{600} (NLS::mScarlet::NLS) in agroinfiltration buffer (10 mm MgCl2, 10 mm MES, pH 5.6). *N. benthamiana* plants were grown and maintained in a temperature-controlled growth room at 25 °C and 60% humidity in 16:8 h light:dark cycles with a daytime PPFD of ~120 µmol/m2 s. *N. benthamiana* were germinated and grown in Sungro Sunshine Mix #4 supplemented with ICL Osmocote 14-14-14 fertilizer at 5 mL/L and agroinfiltrated at 29 days of age. Constructs of interest were infiltrated into the fourth leaf (counting down from the top of the tobacco plant) and harvested 3-4 days post-infiltration. Images were taken using a Zeiss LSM710 laser-scanning confocal microscope at 630x magnification.

4.3.8 Preparation of *Mesorhizobium loti* for Plant Inoculation

M. loti strains were grown for 48hrs at 28°C, 180 rpm in TY media. The cultures were then transferred to 15 or 50 mL labeled falcon tubes and pelleted using a swing arm centrifuge at 4000 rcf for 10-15 min. The media was decanted and bacteria were resuspended in DI H2O before two rounds of washing. The bacteria were resuspended a final time in DI H20 and adjusted to the desired OD_{600}.

4.3.9 Destructive Sampling Sand Cone Experimental Setup

Lotus japonicus seedlings were germinated as described and at 7 days old transplanted to 4.75-inch cone-tainers (Ray Leach Stuewe and Sons) filled with 0.3 cm^3 rockwool at the base, then 27.5 mL medium grain sand (Cemex Lapis Lustre Specialty Sands NO#60), and topped with 10 mL fine play sand (SAKRETE). For uninoculated experiments, plants were watered with a 1/2 MS nutrient solution, covered with cling wrap and a humidifier dome, then placed in the growth chamber. They remained covered for 1 week and were then watered with 1/2 MS once a week for the duration of the experiment. Samples were harvested at desired time points by hand, roots were cleaned with DI water, patted dry with a paper towel, then weighed and images were taken for later measurement in ImageJ.

For inoculated experiments, plants were watered with a 1/2 MS without nitrogen solution, covered with cling wrap and a humidifier dome, then placed in the growth chamber. They remained covered for 1 week before flood inoculation with 3 mL of 0.1 OD_{600} of the desired *M. loti* strain per plant and were then watered once a week with 1/2 MS without nitrogen for the duration of the experiment. Following inoculation, samples were harvested at desired time points by hand, roots were cleaned with DI water and assessed for (im)mature nodule formation while remaining in water to prevent desiccation using an Olympus SZX16 stereoscope fitted with an Olympus SZX2-ILLT

base for transmitted/oblique illumination, then patted dry with a paper towel. Finally, they were weighed, and images were taken for later measurement in ImageJ.

4.3.10 Non-destructive Sampling Nodule Plate Experimental Setup

50 mL slanted agar plates are 1/4 B&D (Broughton and Dilworth) nutrient media without nitrogen, 1.5% w/v agar (PhytoTech Labs A111). Autoclaved plant germination paper cut to fit is wet with the same nutrient solution and placed over the agar. Autoclaved metal bars to hold the seedlings and shade roots from light are then placed into the plates. Seedlings were transferred ~4-7 days post germination then plates were wrapped 3/4 with parafilm with 1/4 Millipore tape (3M) at the top, placed in a black cardstock cover to shield roots from light, and left at the described growing conditions. One week after transplanting, each plate was inoculated with 500 mL 0.05 OD_{600} *M. loti* RFP, re-wrapped, and returned to the growth chamber. Images were taken every 2-dpi using the Amersham Imager 600 (GE) blue and green light channels for 0.5 seconds each. Additional color photos were taken with a standard camera. Images were analyzed in ImageJ to quantify nodule formation and maturation.

4.3.11 Data Analysis

All quantitative data analysis and visualization occurred in Jupyter notebook with custom Python scripts using the SciPy, NumPy, Pandas, Matplotlib, and Seaborn libraries. The nonparametric Kruskal Wallis test was used for all nodulation experiment data.

4.4 Results

4.4.1 Symbiosis-responsive SXLGs share conserved features of the XLG N-terminal domain

Figure 4-1. SXLGs share conserved features of the XLG N-terminal domain.

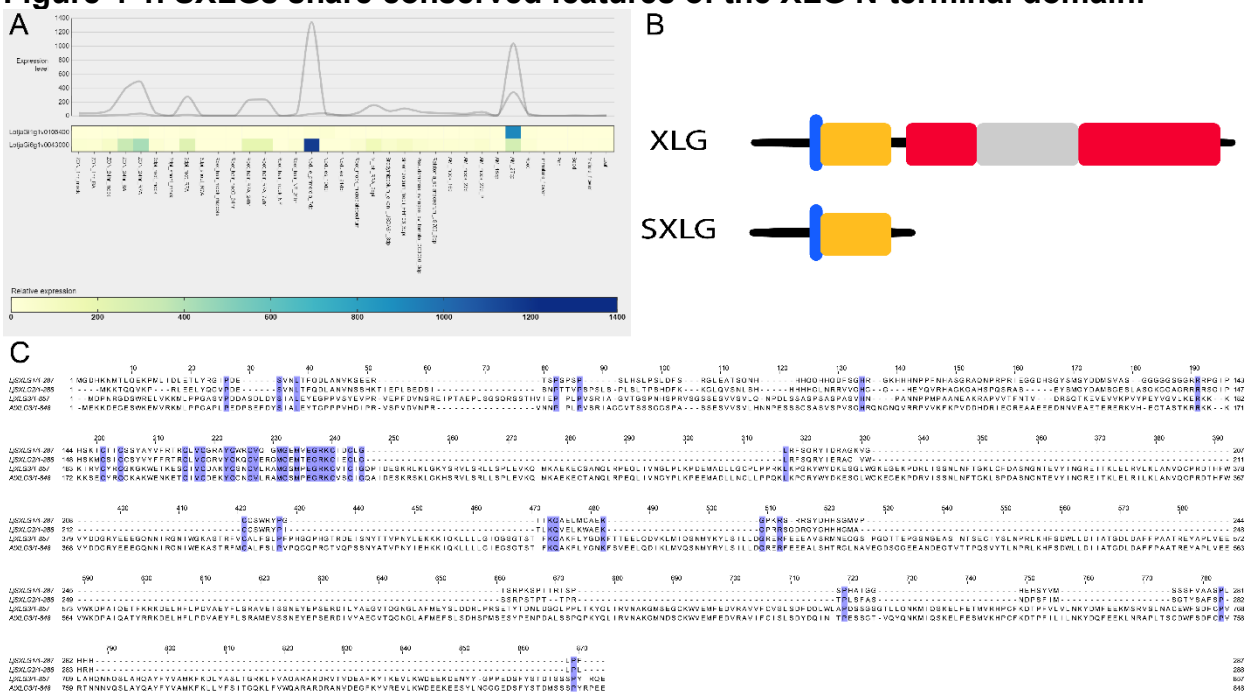


Figure 4.1. A) LotusBase Expression Atlas line and heat map displaying the measured expression levels of *LjSXLG1*(LotjaGi1g1v0106400) and *LjSXLG2*(LotjaGi6g1v0043000) across tissues and conditions. Note the tightly controlled patterns: *SXLG1* at 28dpi mycorrhizal conditions and *SXLG2* also during early stages of the rhizobial symbiosis (recognition and immature nodules) **B)** Graphical representation of XLG domains and the areas SXLGs show the highest conservation, blue = NLS, gold = cysteine-rich domain, red = RAS-like domain, gray = helical domain. Red and gray constitute the Gα-like domain **C)** Muscle peptide sequence alignment of *Lj*SXLGs, *Lj*XLG3, and *At*XLG3. Fully conserved amino acids are highlighted purple, clearly indicating the shared cysteine-rich domain.

Utilizing the peptide BLAST and Expression Atlas features of LotusBase**,** we identified two homologs of *Lotus japonicus* XLG3 with transcriptional expression profiles associated with the establishment of mycorrhizal and root nodule symbioses ,SXLG1(LotjaGi1g1v0106400) and SXLG2(LotjaGi6g1v0043000) respectively. Multiple sequence alignment of SXLG and XLG peptides revealed the conservation to be located in the largely uncharacterized N-terminal domain of conventional XLGs and that SXLGs were approximately one-third their size, the sub-domain of highest conservation contained the putative NLS and the cysteine-rich DUF characteristic of XLGs (Figure 4-1).(517) This highly conserved DUF has the InterPro annotation “Zinc Finger, RING/FYVE/PHD-type”, sharing the closest resemblance to the Plant HomeoDomain (PHD) motif but lacking the histidine pattern**.**(535)

As the size of SXLGs were comparable to the other classes of G-proteins in plants, analysis of InterPro domain annotation in Phytozome and multiple sequence alignments were used to see if they were indeed a class of XLGs. SXLGs did not share significant alignment with the defining domains of other heterotrimeric or small monomeric G-proteins (Supplemental Figure 4-1).

4.4.2 Phylogenetic analysis indicates duplication and neofunctionalization of SXLGs in *Fabaceae* species

Figure 4-2. Phylogenetic analysis indicates duplication and neofunctionalization of SXLGs in *Fabaceae* species

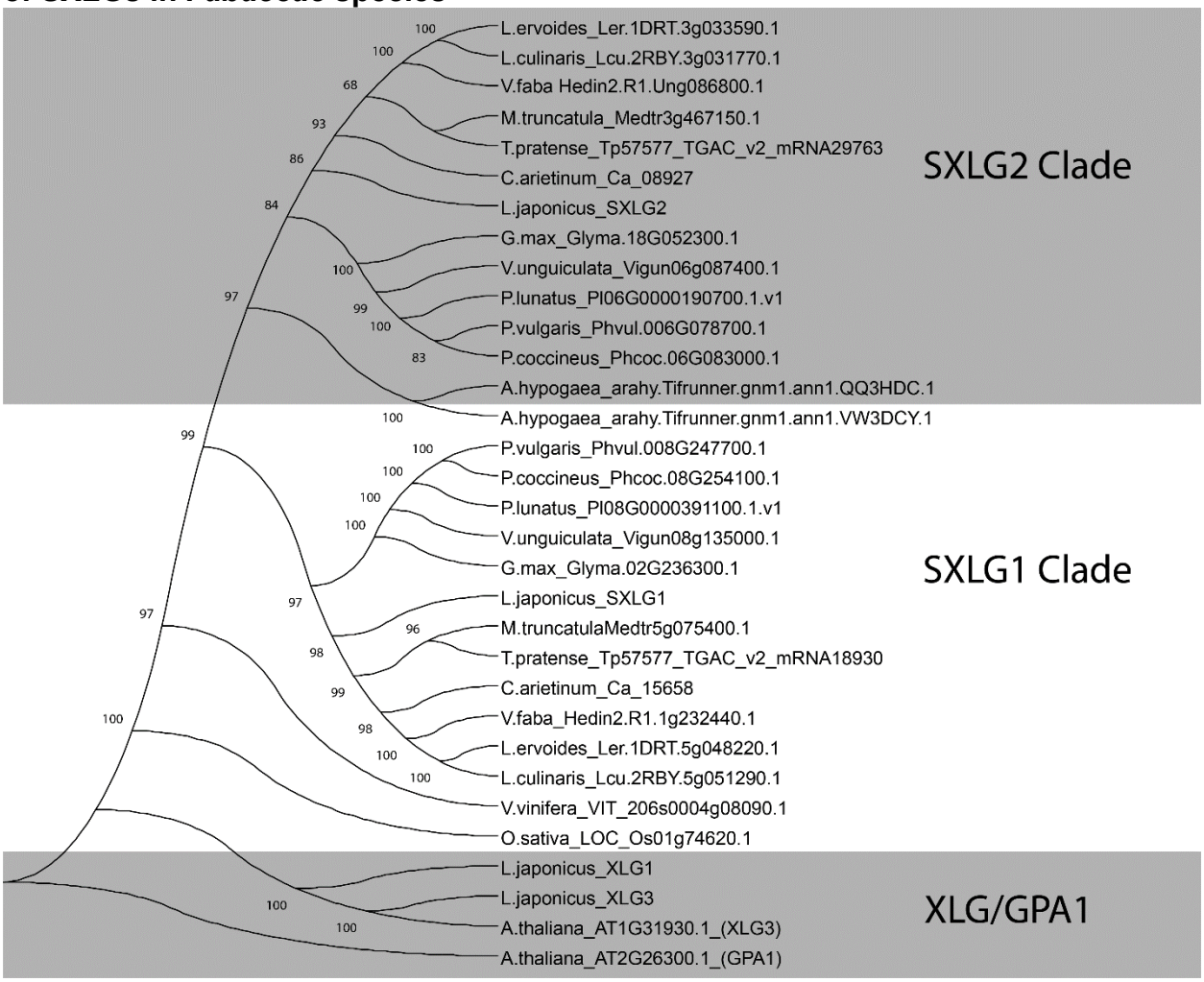


Figure 4-2. Phylogenetic tree showing SXLGs separate into two clades in *Fabaceae* and away from GPA1 and canonical XLGs. The SXLG1 clade contains non-*Fabaceae* SXLGs and expression data in *L. japonicus* and *M. truncatula* show members to be primarily responsive to mycorrhizal symbiosis. The SXLG2 clade separates furthest from non-*Fabaceae* and expression data show members to be primarily responsive to rhizobial symbiosis. There also seems to be grouping according to nodule type (determinate and indeterminate).

Many proteins involved with the common symbiotic pathway have undergone duplication and neofunctionalization, with the new copies developing some distinct functions associated with rhizobial symbiosis while older ones retain functions primarily in mycorrhizal symbiosis.(536) As there were two identified SXLGs in *L. japonicus*, each with expression profiles biased towards one symbiosis, it raised the possibility of the genes having undergone this process. Indeed, a phylogenetic analysis indicated such a pattern, with SXLG homologs across *Fabaceae* species forming two distinct clades — one containing the mycorrhizal-responsive SXLG1 and one with the rhizobia-responsive SXLG2 from *L. japonicus* (Figure 4-2). Additionally, transcriptomic data from MtExpress

indicated the *Medicago truncatula* homologs *MtSXLG1*(MtrunA17_Chr5g0432711; Medtr5g075400) and *MtSXLG2*(MtrunA17_Chr3g0107521;Medtr3g467150) grouped according to the same bias in expression profiles.(537) The putative SXLG1 *Fabaceae* clade was more closely related to putative SXLGs found in plants only able to form mycorrhizal symbiosis (Figure 4-2).

Further supporting this possible duplication event, BLAST results and neighbor-joining tree analysis consistently demonstrated the number of SXLG clades was two in leguminous *Fabaceae* plants and one in plants only able to form mycorrhizal symbiosis (data not shown). Interestingly, *Brassicaceae* species — which are unable to form either symbioses — were found to lack SXLGs entirely.

4.4.3 Transient expression reveals cellular localization matching conventional XLGs

Figure 4-3. Transient expression reveals cellular localization to plasma membrane, nucleus, and nucleolus

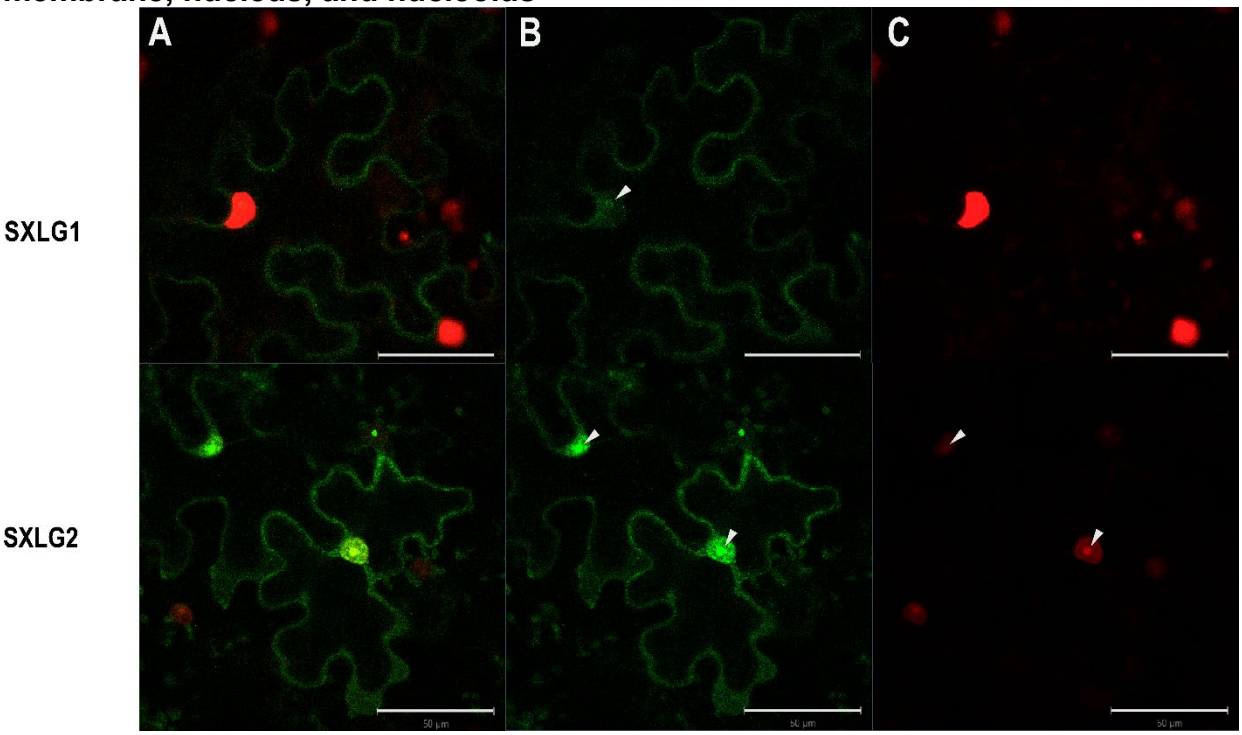


Figure 4-3. Cells are transfected with a plasmid containing mNeonGreen (mNG) fused to the C-terminus of the coding sequence for *LjSXLG1* (Top) *or LjSXLG2* (Bottom), driven by the pCM2 promoter. An additional plasmid containing mScarlet flanked by a nuclear localization signal is utilized as a nuclear marker. Results were consistent across promoter strength, infiltration OD, and with a N-terminal mNG fusion **A)** Overlay of **B)** LjSXLG::mNeonGreen fusion and **C)** nuclear marker images. LjSXLGs are localized to the plasma membrane, nucleus, and nucleolus. Nucleoli are marked by white arrows. Bars = 50uM. All images are 630x magnification.

XLGs have demonstrated the ability to localize to both the plasma membrane and nucleus, indicating they are not membrane bound like conventional Gα subunits and can traffic through the cytoplasm.(538) Transient expression of *LjSXLG1* and *LjSXLG2* - fused to mNeonGreen in *N. benthamiana* leaves revealed the same capability, with

SXLG1 exhibiting reduced nuclear localization compared to SXLG2 (Figure 4-3). However, both were able to be imported into the nucleolus, something not yet demonstrated in XLGs and absent in our free mNeonGreen control agroinfiltration (Supplemental Figure 4-2).(517) These results persisted across a range of promoter strengths and with mNeonGreen fused to either terminus.

4.4.4 Destructive sampling reveals transient impairment of immature nodule formation in *Lotus japonicus sxlg2* mutants

Figure 4-4. *Lotus japonicus sxlg2* mutants have transient impairment of immature and mature nodule formation

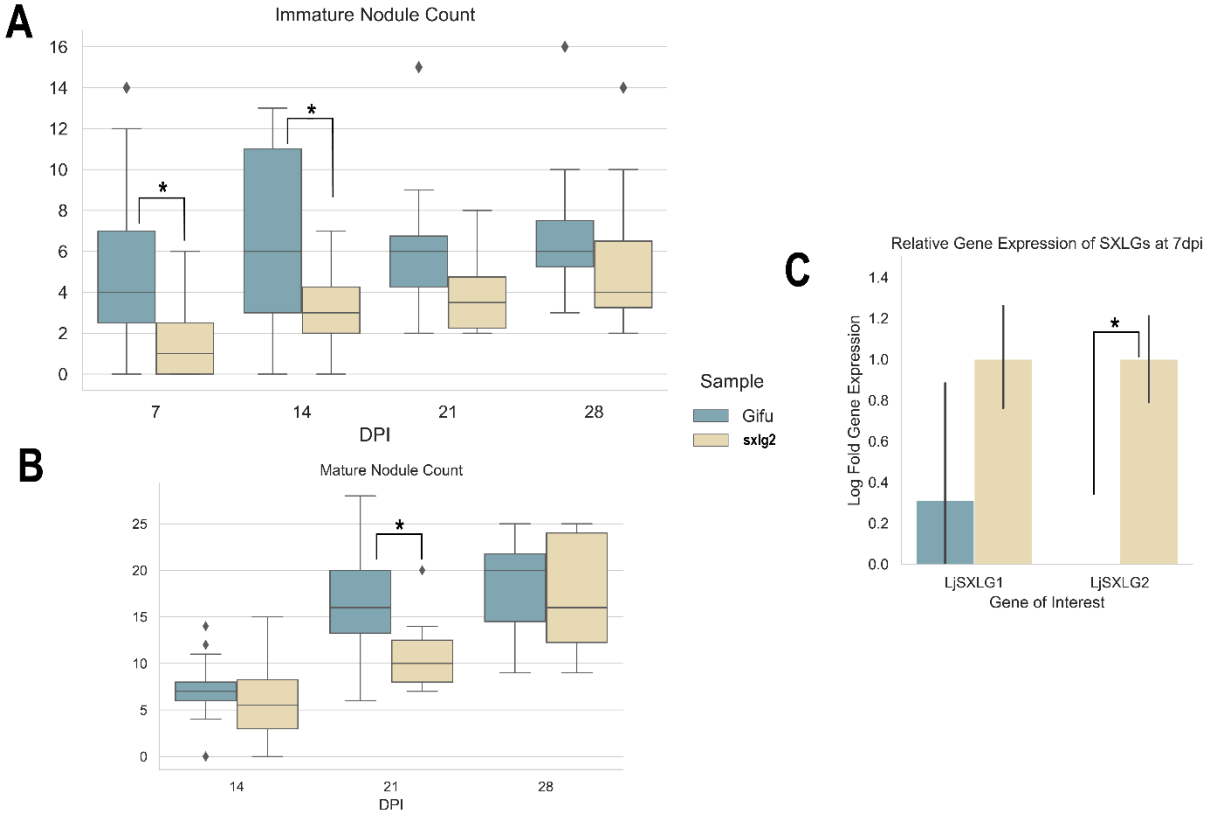


Figure 4-4. Comparison of WT Gifu & *Ljsxlg2* plants grown in nitrogen depleted conditions in sand cones. **A)** Immature nodule & **B)** Mature nodule count n = 19 (7dpi), 25 (14dpi), & 10 (21 and 28dpi) **C)** Expression levels of *SXLG1* and *SXLG2* using *LjUBQ*(LotjaGi5g1v0317900) as reference. n = 3. P-value ≦ 0.02 denoted by *. Calculated by Kruskal-Willis Test.

LotusBase ExpressionAtlas shows *LjSXLG2* (LotjaGi6g1v0043000) is primarily expressed in immature nodules and to a lesser extent in early stages of association with rhizobia.(503) Wildtype Gifu and transposon knockout mutant plants were grown in sand cones and inoculated with *M. loti* R7A to assess potential effects to nodulation. At 7- and 14-days post inoculation (dpi) *sxlg2* mutants showed less immature nodules ($P \leq 0.02$) and maintained a depressed, but statistically insignificant, number through the remainder of the experiment compared to Gifu (Figure 4-4A). However, this only translated into a reduction of mature nodules at 21 dpi ($P \leq 0.02$) (Figure 4-4B). There were no observed differences in the size, coloration, or structure of nodules (data not shown).

4.4.5 Non-destructive observations of *Lotus japonicus sxlg2* mutants show reduced number of established infection events with no effect on nodule maturation

Figure 4-5. *Lotus japonicus sxlg2* mutants show reduced number of established infection events with no effect on nodule maturation

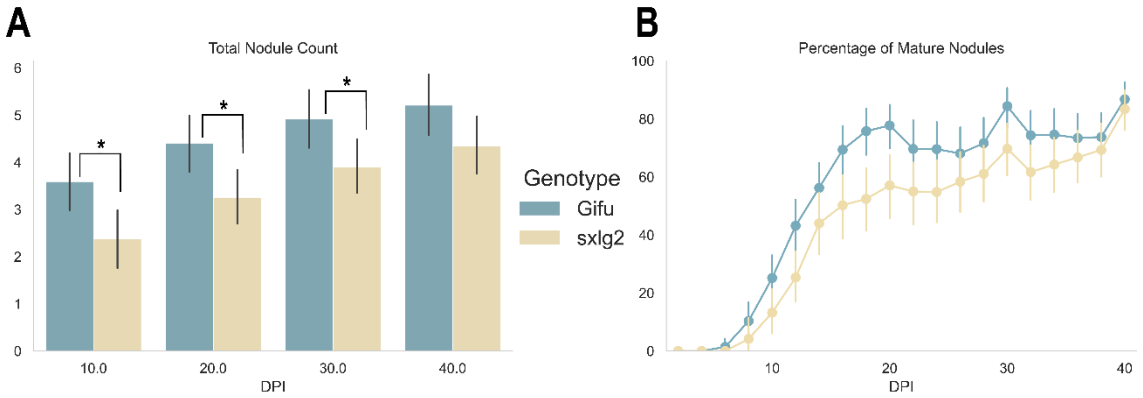


Figure 4-5. A) Comparison of total nodule counts between WT Gifu & Lj*sxlg2* plants grown in nitrogen depleted conditions on plates that allowed for non-destructive sample collection. **B)** The percentage of nodules that were mature at time of sampling. P-value ≦ 0.02 denoted by *. Calculated by Kruskal-Willis Test. n = 70

Considering the transient nature of the immature nodule phenotype and its inconsistent translation into a reduction of mature nodules, we followed up with non-destructive observations to see if destructive sampling was obscuring important dynamics of the symbiotic infections. Gifu and *sxlg2* mutants were grown on low-nutrient, nitrogen-depleted agar plates and inoculated with *M. loti* expressing red fluorescent protein to enable continuous observation. The limited nutrient content also restricted the plants to approximately 2-3 waves of established infection events (where nodules appeared). Through 30 dpi *sxlg2* mutants formed less established infection events than Gifu ($P <$ 0.02), and when distinguishing mature nodules the reduction was maintained through 40 dpi (Figure 4-5A, Supplemental Figure 4-3). The ratio of immature:mature nodules was measured at 2 day increments to determine if nodules in *sxlg2* mutants had altered maturation rates. Little to no differences were observed for the duration of the experiment (Figure 4-5B).

4.4.6 *Lotus japonicus sxlg2* mutants exhibit potential impairment of root growth response to nitrogen limitation

Figure 4-6. *Lotus japonicus sxlg2* mutants exhibit potential impairment of root growth response to nitrogen limitation

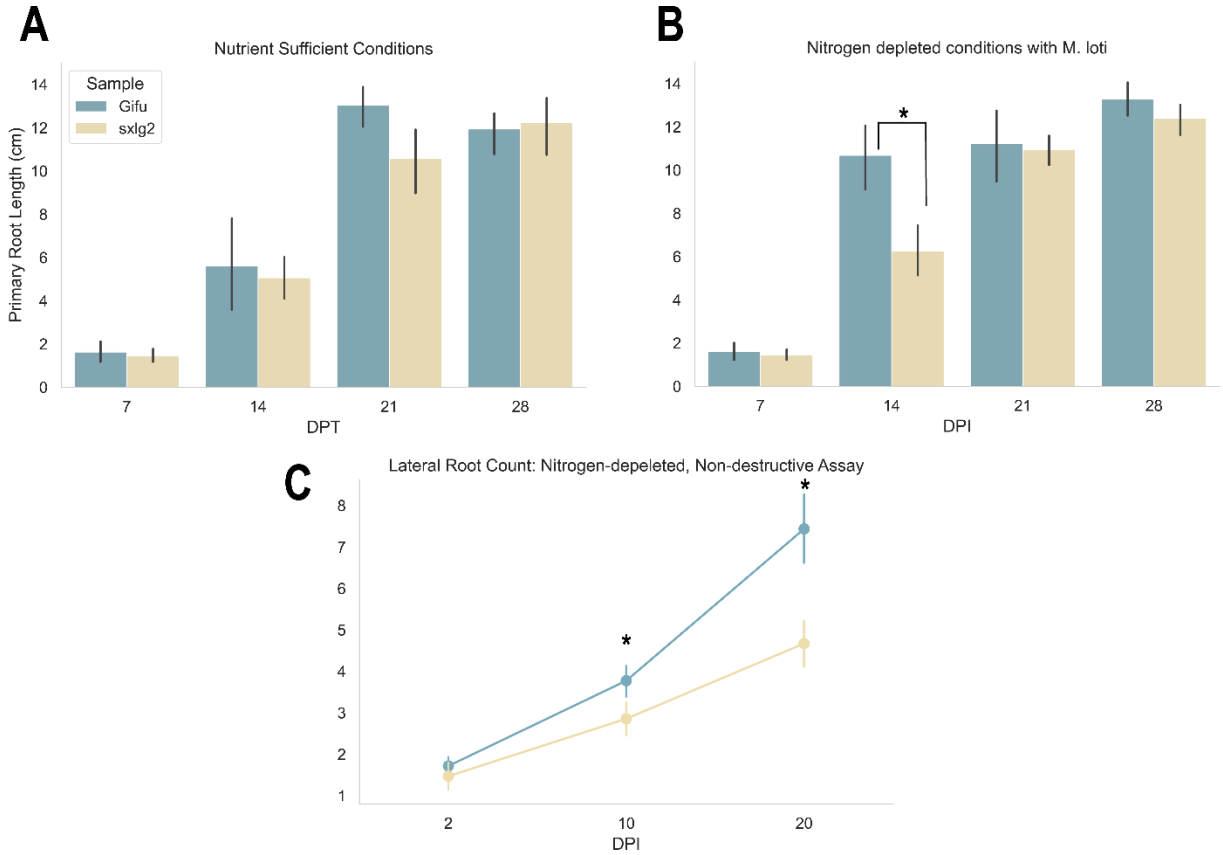


Figure 4-6. Comparison of primary root length between WT Gifu & Lj*sxlg2* plants grown in **A)** Nutrient sufficient, uninoculated conditions and **B)** Nitrogen depleted, *M. loti* inoculated conditions in sand cones. n = 8-10. **B)** The percentage of nodules that were mature at time of sampling. **C)** Lateral root count through 20 dpi in the nitrogen-depleted, non-destructive conditions on plates inoculated with *M.* loti. n = 70. P-value ≦ 0.02 denoted by *. Calculated by Kruskal-Willis Test.

Plants will preferentially allocate resources to grow their root systems when facing nitrogen limitation.(539,540) Physiological root data in both the destructive and non-destructive experimental settings indicate a potential impairment of the exploration response to nitrogen starvation. In the sand cone-tainer nitrogen-depleted experiments primary root length in *sxlg2* mutants did not show the increase at 14 dpi compared to nutrient sufficient conditions that was seen in Gifu (Figure 4-6A,B). Additionally, in the nitrogen-depleted non-destructive experiments *sxlg2* mutants had reduced lateral root branching (Figure 4-6C).

4.5 Discussion

4.5.1 SXLGs may provide insight into the evolutionary history and molecular functions of XLGs

SXLGS share the strong conservation of the cysteine DUF seen across XLGs and do not seem to share conserved features found in other classes of plant G-proteins,

whether mono- or heterotrimeric, indicating they may be a novel subclass of XLGs (Figure 4-1, Supplemental Figure 4-1).(515,516,541) Additionally, similar to XLGs being a class of proteins unique to plants, pBLAST results for both *Lj*SXLGs across the UniProtKB/Swiss-Prot and non-redundant protein sequences databases only provided potential homologs among terrestrial plants and some algae.(542) Nearly all of these were either characterized XLGs, annotated as putative XLGs or XLG-like, or uncharacterized proteins. Entering only the cysteine-rich DUF provided the same results, with the exception of three hits; oomycetes and fungi that are either pathogenic or symbiotic to plants. Further, the potential diversity of SXLG homologs (one to two copies per diploid genome) more closely resembles heterotrimeric G-protein subunit diversity (one to four copies per diploid genome) than the typical wide diversity of monomeric G-proteins.(515,543)

Considering that there is currently no known function of this highly conserved cysteine-rich domain and it is believed XLGs arose in plants due to rapid evolutionary steps, such as a potential fusion of the C-terminal Gα-like domain with the uncharacterized N-terminal domain, SXLGs present a missing piece of the XLG puzzle.(544) While the methods used in this study cannot accurately determine if SXLGs are a candidate for the protein which fused to the Gα-like domain or one that arose from separation of the N- and C-terminal domains, it is clear they contain the to-date uncharacterized domain found in all XLGs (Figure 4-1B)**.** Further study on SXLG molecular structure and function could yield critically lacking information about XLGs.

4.5.2 SXLGs may be involved in transcriptional regulation

This study demonstrated SXLGs possess a functional NLS (Figure 4-3). They also share some alignment with the nuclear export signal in XLGs, however potentially critical hydrophobic residues are lacking and further work is needed to confirm its functionality.(517) Interestingly, both *Lj*SXLGs were shown to localize to the nucleolus in our investigations, which has yet to be demonstrated in XLGs and was absent in our free mNeonGreen control (Figure 4-3, Supplemental Figure 4-2).(517)The nucleolus is a site where various regulatory processes occur, such as those involved in the cell cycle and pre-mRNA splicing.(545)The cysteine-rich DUF characteristic of XLGs is annotated as a zinc finger domain and, excluding the histidine residues, shares close resemblance to the PHD motif, which is commonly associated with chromatin modifiers and various transcription factors in plants.(535,546)

The ability to be present at the plasma membrane and in the nucleus and nucleolus, alongside the cysteine-rich domain's resemblance to the PHD motif, is intriguing evidence of SXLGs potential to contribute to transcriptional regulation in some form. Lending to this hypothesis, XLG2 has been shown to be involved in transcriptional regulation via interaction with RELATED TO VERNALIZATION1 (RTV1) in the nucleus to stimulate chromatin binding.(547) Further study of SXLGs could reveal if they are indeed involved in transcriptional regulation and shed light on the molecular mechanisms other XLGs may be utilizing. This is currently complicated by the lack of a defined protein structure for XLGs beyond the Gα-like domain, attempts to generate structures of SXLGs or XLGs *in silico* result in very low-confidence scores outside of a

portion of the cysteine-rich domain, which is predicted to form an α-helix (data not shown). Additionally, if SXLGs share the ability to bind to GTP/GDP, Gβ**γ** dimers, or undergo phosphorylation like XLGs is currently unknown, however XLG2 was shown to be phosphorylated in the N-terminal domain at serine residues upstream the cysteine DUF which *Lj*SXLG1/2 also possess (Figure 4-1B).(517,548)

4.5.3 SXLGs are likely involved in the establishment of rhizobial and mycorrhizal symbioses

Likely due to their roles in organ development and cell proliferation, many G-proteins have been demonstrated to be involved in various stages of rhizobial and mycorrhizal symbiosis.(518,525) All heterotrimeric G-protein subunits are responsive to or directly involved during the symbioses: Gα were shown to directly interact with NFR1a and NFR1b, Gβ expression levels impact nodule formation, G**γ** transcription is up-regulated in both nodulation and mycorrhizal symbiosis, and regulator of G-protein signaling (RGS) expression levels impact nodule formation.(519–521,549) There is also increasing evidence that monomeric small GTPases of the ROP family are influential contributors to infection thread (IT) development.(550–552) However, there has been little direct evidence of XLGs involvement in either symbiosis.

A combination of phylogenetic analysis, available transcriptomic data, and experimental observations has indicated that SXLGs may be involved in the establishment of rhizobial and mycorrhizal symbiosis. The potential duplication and neofunctionalization of SXLGs seen in *Fabaceae*, as well as their absence in *Brassicaceae*, indicates involvement in symbiotic relationships (Figure 4-2). Data from LotusBase and Frank et al. 2023 have shown *L. japonicus SXLG2* to be expressed during early recognition of compatible rhizobia, in immature nodules, and moved from root hair to cortical cells within the first 10 dpi with rhizobia (Figure 4-1A).(503,553) Similarly, data from multiple studies in MtExpress showed *MtSXLG2*(MtrunA17_Chr3g0107521;Medtr3g467150) to be expressed during nitrogen starvation, early recognition of compatible rhizobia, ectopic addition of Nod factors, and nodule Zones I and II — where ITs remain present. In Serrano et al. 2024 and MTExpress *MtSXLG1* (MtrunA17_Chr5g043271; Medtr5g075400) and *MtSXLG2* were shown to be expressed in cortical cells during stages of mycorrhizal association when arbuscules are present.(537,554–560)

These are stages of the symbioses where Nod factor recognition is ongoing and the IT and peri-arbuscular membrane (PAM) are created and maintained.(561–563) Our experiments demonstrated *L. japonicus sxlg2* mutants to be initially impaired in developing established infection events with the compatible rhizobia *M. loti* (Figure 4-4, Figure 4-5). However, the nature and maturation rates of the nodules did not change, implying SXLGs are uninvolved in developmental processes once nodules have been established. The involvement of small GTPases in IT development, mature determinate nodules of *L. japonicus* not maintaining ITs, and *MtSXLG2* expression only in zones of indeterminate nodules were ITs are present all contribute to the possible role of SXLGs in IT development. *SXLG1* expression patterns also indicate potential involvement in PAM development or maintenance. However, further work is required to confirm if SXLGs are involved in these processes and the nature of their participation.

4.6 Conclusion

Our study has introduced a novel sub-class of XLGs with conserved features of the N-terminal domain and lacking a Gα-like domain. The cysteine-rich domain within the N-terminus may be involved in transcriptional regulation based on similarity to the PHD domain and SXLG and XLG cellular localization patterns. SXLGs appear to be involved in symbiotic relationships; indication of duplication and neofunctionalization in *Fabaceae*, available transcriptomic data in *L. japonicus* and *M. truncatula*, and experiments showing *LjSXLG2* influences the establishment of infection events during rhizobial symbiosis contribute to this hypothesis. SXLGs present an ideal opportunity to better understand the evolution, function, and structure of XLGs and are another example of G-proteins involvement in symbiotic relationships.

4.7 Acknowledgements

This work was not possible without the assistance of many people. Tomo Yoshino, Thalissa Malagoli Franzon, and Victoria Vera assisted in conducting experiments, collecting data, and data analysis. Dr. Mitchell Thompson, Dr. Andy Zhou, Dr. Chris Gee, Dr. Simon Alamos, and Liam Kirkpatrick all contributed essential materials and technical support for conducting experiments. Dr. Henrik Scheller assisted in experiment design, data analysis, and manuscript editing. This work was supported by the InRoot Project, which is funded by the NovoNordisk Foundation, & conducted at the Joint BioEnergy Institute which is supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Lorenzo Washington was further supported by the National Science Foundation under Grant No. DGE 2146752 & the Berkeley Fellowship for Graduate Study.

4.8 Supplemental Information

Located in Appendices

Supplemental Table 4-1. Primers used in study

Supplemental Table 4-2. Raw log-fold ddCt values for qPCR

Supplemental Figure 4-1. Alignments of SXLGs to other G-protein subunits and collective phylogenetic tree

Supplemental Figure 4-2. Transient Expression of free mNeonGreen and N-terminal fusion of SXLG2

Supplemental Figure 4-3. Non-destructive observations separated by (im)mature nodule status

Chapter 5. Summary and Future Directions

Agriculture is one of humanity's earliest and most disruptive applications of plant biology research. It paved the way for the growth of human population and development of large, complex societies which began a feedback loop of curiosity and innovation that is still felt today. However, history is riddled with nuance and unexpected outcomes; developments frequently raise as many or more issues than they address through revealing context we didn't even know we didn't know about. As era-defining challenges intensify, it is necessary for us to be critical of the historical and contemporary contexts and material realities that shape the opportunities we have to problem solve and deepen our understanding of the world. Only then will we be situated to make well-informed decisions as the climate crisis, ecological, and environmental degradation accelerate. This dissertation is an attempt to practice that criticality while contributing to our fundamental understanding of the connectedness within, between, and beyond the various categories humans use while trying to make sense of the world.

Chapter 1 delved into the historical and contemporary influences of Western imperialism, colonialism, and capitalism on agriculture and plant biology through the institution of Science. It detailed a 500+ year thread of socioeconomic power that continues to leave its mark on how we think about and practice scientific research. The understanding that the globalized industrial agricultural system we rely on today is a direct evolution of plantation agriculture, which prioritized the production of cash crops over meeting local nutritional needs, environmental stewardship, and human well-being, allows us to assess with greater clarity the materials, practices, and technologies which are available to work with as well as what directions we need to take if we wish to depart from systems perpetuating pressing global challenges.

Especially in the decades following the Green Revolution, we face both a lack of information on root biology and its ecological and environmental intersections and a substantial reduction in the diversity of resources for study of and application towards alternative styles of agricultural production. Chapter 2 detailed and demonstrated the importance of plant roots to the broader ecosystem, hurdles to advancing our understanding, and how recent innovations can be utilized to further accelerate the growth of that understanding. Many difficulties remain, but much progress has been forged in the last two decades.

Chapter 3 illuminated underlying mechanisms of pectin biosynthesis through demonstrating RRTs involvement in RG-I biosynthesis in root tissue of *Lotus japonicus*. Cell wall biology remains a complicated and arduous area of study due to considerable functional redundancy and complexity in the dynamics between cell wall biosynthesis and a variety of cellular signaling pathways. However, investigations continue to provide useful insights for application to breeding and engineering plants and plant roots well-suited for novel agricultural approaches and valuable product and biofuel production.

Chapter 4 characterized SXLGs, a previously unknown subclass of plant GTPase-related signaling proteins which appear to influence symbiotic relationships formed in root systems. They share the uncharacterized N-terminal domain of XLGs but lack the

Gα-like C-terminal domain. Phylogenetic and molecular genetic approaches indicate SXLGs are involved in symbiotic relationships exclusive to root systems. These represent an ideal opportunity to elucidate the evolution and function of the uncharacterized N-terminal domain and cysteine-rich motif present in all identified XLGs. There may be potential applications in breeding and engineering plants specifically for their abilities to form symbioses, as well as synthetic biology applications if the cysteine motif is demonstrated to be involved in transcriptional regulation.

In its own way, each of these chapters played a part in unearthing another strand of information that can help us understand the larger picture. Whether they're building the economic power of globe-spanning empires, playing essential roles in nutrient cycling and soil health, shuffling sugars around to build sturdy and dynamic scaffolds to support themselves, navigating the difficulties of intimate relationships in a shared space, or helping to put food on the table and oxygen in the air — From many perspectives, plant roots hold a special place in this world and provide support for far more than the plant they are growing from. That alone warrants considerable attention and could hold solutions to problems we have yet to be aware of. It has been both frustrating and a privilege to peek into the inner workings of such fundamentally important creatures. I hope my work helps others to uncover and share their own sliver of knowledge and encourages being as critical of the contexts around our work as we are of the work itself.

References

1. Blatch S. Great achievements in science and technology in ancient Africa [Internet]. 2013 [cited 2024 May 7]. Available from: https://www.asbmb.org/asbmb-today/science/020113/great-achievements-in-stem-in-ancient-africa

2. Faruqi Y. Contributions of Islamic scholars to the scientific enterprise . International Education Journal: Comparative Perspectives. 2006;

3. Kettani MA. Moslem contributions to the natural sciences - UNESCO Digital Library [Internet]. 1976 [cited 2024 May 6]. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000020204

4. Yates C. Five ways ancient India changed the world – with maths [Internet]. 2017 [cited 2024 May 6]. Available from: https://theconversation.com/five-ways-ancient-india-changed-the-world-with-maths-84332

5. Carpi A, Egger A. A brief history of scientific practice Visionlearning Vol. HID (1) [Internet]. 2009 [cited 2024 Jun 11]. Available from: https://www.visionlearning.com/en/library/Astronomy/5//164

6. Syfret RH. The origins of the Royal Society. Notes Rec R Soc Lond. 1948 Apr 30;5(2):75–137.

7. Patin B, Sebastian M, Yeon J, Bertolini D, Grimm A. Interrupting epistemicide: A practical framework for naming, identifying, and ending epistemic injustice in the information professions. J Assoc Inf Sci Technol. 2021 Apr 7;

8. Adas M. Colonialism and Science. In: Selin H, editor. Encyclopaedia of the History of Science, Technology, and Medicine in Non-Western Cultures. Dordrecht: Springer Netherlands; 2008. p. 604–9.

9. Naithani S. History and Science of Cultivated Plants. Oregon State University; 2021.

10. Liboiron M. Pollution Is Colonialism. Duke University Press; 2021.

11. Egerton FN. A History of the Ecological Sciences, Part 10: Botany during the Italian Renaissance and Beginnings of the Scientific Revolution. The Bulletin of the Ecological Society of America. 2003 Jul 1;

12. LaHaye L. Mercantilism. In: Palgrave Macmillan, editor. The new palgrave dictionary of economics. London: Palgrave Macmillan UK; 2017. p. 1–4.

13. McMahon T. Origins of White Supremacy: The Only ‘Doctrine’ That Actually Matters. SSRN Journal. 2022;

14. Pope Alexander VI. Inter caetera by Pope Alexander VI [Internet]. Encyclopedia Virginia, Virginia Humanities. 1493 [cited 2024 Apr 23]. Available from: https://encyclopediavirginia.org/primary-documents/inter-caetera-by-pope-alexander-vi-may-4-1493/

15. Wolf K. The “Moors” of West Africa and the Beginnings of the Portuguese Slave Trade. Journal of Medieval and Renaissance Studies. 1994;24:457–9.

16. Malm A. Fossil Capital: The Rise of Steam Power and the Roots of Global Warming. Illustrated. London: Verso; 2016.

17. Harrison M. Science and the british empire. Isis. 2005 Mar;96(1):56–63.

18. Baber Z. The plants of empire: botanic gardens, colonial power and botanical knowledge. J Contemp Asia. 2016 Oct;46(4):659–79.

19. Schiebinger L. Plants and empire: colonial bioprospecting in the atlantic world. Harvard University Press; 2009.

20. Bala A. The Dialogue of Civilizations in the Birth of Modern Science. 1st ed. 2006.

21. McNeill JR. Columbian Exchange | Diseases, Animals, & Plants | Britannica [Internet]. 2019 [cited 2024 Apr 25]. Available from: https://www.britannica.com/event/Columbian-exchange#ref343810

22. Moore S, Allewaert M, Gomez P, Mitman G. The Plantationocene and Plantation Legacies Today - Edge Effects [Internet]. 2019 [cited 2024 Apr 18]. Available from: https://edgeeffects.net/plantation-legacies-plantationocene/

23. Hancock J. Sugar & the Rise of the Plantation System - World History Encyclopedia [Internet]. 2021 [cited 2024 May 10]. Available from: https://www.worldhistory.org/article/1784/sugar--the-rise-of-the-plantation-system/

24. Sheridan R. Sugar and Slavery: An Economic History of the British West Indies, 1623-1775. Johns Hopkins University Press; 1974.

25. Mapping History : The Spread of Cotton and of Slavery 1790-1860 - The Economics of Cotton [Internet]. [cited 2024 Apr 30]. Available from: https://mappinghistory.uoregon.edu/english/US/US18-04.html

26. Hancock J. Slavery in Plantation Agriculture - World History Encyclopedia [Internet]. 2021 [cited 2024 Apr 30]. Available from: https://www.worldhistory.org/article/1837/slavery-in-plantation-agriculture/

27. Johnson T. Growth industry: fertilizer and the politics of agriculture on the Georgia cotton belt, 1840-1900 [Internet] [Master thesis]. 2010 [cited 2024 Jun 10]. Available from: https://esploro.libs.uga.edu/esploro/outputs/graduate/Growth-industry-fertilizer-and-the-politics/9949334540902959

28. Snively G, Williams WL. Knowing Home: Braiding Indigenous Science with Western Science. 1st ed. University of Victoria; 2016.

29. Kean S. Historians expose early scientists' debt to the slave trade. Science. 2019 Apr 4;

30. Scott-Jones G, CAADC, Kamara MR, PE. The traumatic impact of structural racism on african americans. Dela J Public Health. 2020 Nov 7;6(5):80–2.

31. Beckelheimer T. African Identity and the African Diaspora: The Genetic Impact of the Transatlantic Slave Trade. University of Texas at San Antonio, College of Liberal and Fine Arts. 2021;

32. Fynn Bruey V, Crawley H. The enduring impacts of slavery: A historical perspective on south–south migration. In: Crawley H, Teye JK, editors. The palgrave handbook of south–south migration and inequality. Cham: Springer International Publishing; 2024. p. 25–46.

33. The Legacy of Slavery in the Caribbean and the Journey Towards Justice | United Nations [Internet]. [cited 2024 Jul 27]. Available from: https://www.un.org/en/un-chronicle/legacy-slavery-caribbean-and-journey-towards-justice

34. Brumfield S. The Lasting Impact of Genocide in America: Historical Trauma, Continued Genocide, and Methods of Prevention. University of Nebraska Omaha. 2020;

35. Solomon TGA, Starks RRB, Attakai A, Molina F, Cordova-Marks F, Kahn-John M, et al. The generational impact of racism on health: voices from american indian communities. Health Aff (Millwood). 2022 Feb;41(2):281–8.

36. Smith DM. Counting the Dead: Estimating the Loss of Life in the Indigenous Holocaust, 1492 - Present . University of Houston-Downtown. 2017;

37. Leigh GJ. The World's Greatest Fix: A History of Nitrogen and Agriculture. 1st ed. Oxford: Oxford University Press; 2004.

38. Williams E. Capitalism and Slavery. 3rd ed. University of North Carolina Press; 2021.

39. Dean W. Rio Claro: A Brazilian Plantation System, 1820-1920. Stanford University Press; 1976.

40. Wright G. The Political Economy of the Cotton South: Households, Markets, and Wealth in the Nineteenth Century. W.W. Norton & Company; 1978.

41. Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE, et al. Ecology. Controlling eutrophication: nitrogen and phosphorus. Science. 2009 Feb 20;323(5917):1014–5.

42. Stoll M. Profit: An Environmental History. 1st ed. Cambridge: Polity; 2022.

43. Prashad V. Everybody Was Kung Fu Fighting: Afro-Asian Connections and the Myth of Cultural Purity. Beacon Press; 2002.

44. Serrano Jr OR. Repackaging Plantation Relations: Green Revolution Technologies, Agriculture, and the Remaking of the Américas. Occasion: Race, Space, Scale. 2015 Aug 31;

45. Patel R. The Long Green Revolution. The Journal of Peasant Studies. 2013 Jan;40(1):1–63.

46. Davies WP. An historical perspective from the Green Revolution to the gene revolution. Nutr Rev. 2003 Jun;61(6 Pt 2):S124-34.

47. Pingali PL. Green revolution: impacts, limits, and the path ahead. Proc Natl Acad Sci USA. 2012 Jul 31;109(31):12302–8.

48. John DA, Babu GR. Lessons from the aftermaths of green revolution on food system and health. Front Sustain Food Syst. 2021 Feb 22;5:644559.

49. Eliazer Nelson ARL, Ravichandran K, Antony U. The impact of the Green Revolution on indigenous crops of India. Journal of Ethnic Foods. 2019 Dec;6(1):8.

50. Gollin D, Hansen CW, Wingender A. Two blades of grass: the impact of the green revolution. Cambridge, MA: National Bureau of Economic Research; 2018 Jun.

51. Church N. Why Our Food is So Dependent on Oil [Internet]. Powerswitch. 2005 [cited 2024 May 6]. Available from: https://web.archive.org/web/20060115133725/http://www.powerswitch.org.uk/portal/index.php?option=content&task=view&id=563

52. Garrett RD, Rausch LL. Green for gold: social and ecological tradeoffs influencing the sustainability of the Brazilian soy industry. The Journal of Peasant Studies. 2016 Mar 3;43(2):461–93.

53. Murphy S. Concentrated Market Power and Agricultural Trade. Ecofair Trade Dialouge. 2006 Aug;1.

54. Lynch J, Cain M, Frame D, Pierrehumbert R. Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO2-Emitting Sectors. Front Sustain Food Syst. 2021 Feb 3;4.

55. Liu X, Zhang X, Huang Y, Chen K, Wang L, Ma J, et al. The direct radiative forcing impact of agriculture-emitted black carbon associated with india's green revolution. Earth's Future. 2021 Jun;9(6).

56. Ritchie H, Roser M. Half of the world's habitable land is used for agriculture [Internet]. OurWorldInData.org. 2019 [cited 2024 Apr 15]. Available from: https://ourworldindata.org/global-land-for-agriculture

57. Givens W, Shaw D, Kruger, Johnson. Roundup Ready® Crops Have Major Positive Impact on Tillage Practices . Weed Technology; 2009.

58. Benbrook CM. Impacts of genetically engineered crops on pesticide use in the U.S. -- the first sixteen years. Environ Sci Eur. 2012;24(1):24.

59. Ray DK, Ramankutty N, Mueller ND, West PC, Foley JA. Recent patterns of crop yield growth and stagnation. Nat Commun. 2012;3:1293.

60. Sosnoskie LM, Culpepper AS. Glyphosate-Resistant Palmer Amaranth (Amaranthus palmeri) Increases Herbicide Use, Tillage, and Hand-Weeding in Georgia Cotton. Weed Sci. 2014 Jun;62(02):393–402.

61. Wilkerson J. Why Roundup Ready Crops Have Lost their Allure - Science in the News [Internet]. 2015 [cited 2024 May 6]. Available from: https://sitn.hms.harvard.edu/flash/2015/roundup-ready-crops/

62. Hubbard K. Out of Hand: Farmers Face the Consequences of a Consolidated Seed Industry. Farmer to Farmer Campaign; 2009 Dec.

63. Hayes TO, Kerska K. PRIMER: Agriculture Subsidies and Their Influence on the Composition of U.S. Food Supply and Consumption - AAF [Internet]. 2021 [cited 2024 May 13]. Available from: https://www.americanactionforum.org/research/primer-agriculture-subsidies-and-their-influence-on-the-composition-of-u-s-food-supply-and-consumption/

64. Ergin M, Alkan A. Academic neo-colonialism in writing practices: Geographic markers in three journals from Japan, Turkey and the US. Geoforum. 2019 Aug;104:259–66.

65. Marks RA, Hotaling S, Frandsen PB, VanBuren R. Representation and participation across 20 years of plant genome sequencing. Nat Plants. 2021 Dec;7(12):1571–8.

66. Dahdouh-Guebas F, Ahimbisibwe J, Van Moll R, Koedam N. Neo-colonial science by the most industrialised upon the least developed countries in peer-reviewed publishing. Springer Science and Business Media LLC. 2003;

67. Boshoff N. Neo-colonialism and research collaboration in Central Africa. Scientometrics. 2009 Nov;81(2):413–34.

68. Shiva V. The Violence of the Green Revolution: Third World Agriculture, Ecology, and Politics (Culture of the Land). Reprint. The University Press of Kentucky; 2016.

69. Dinu M, Whittaker A, Pagliai G, Benedettelli S, Sofi F. Ancient wheat species and human health: Biochemical and clinical implications. J Nutr Biochem. 2018 Feb;52:1–9.

70. Valente J, Gerin F, Le Gouis J, Moënne-Loccoz Y, Prigent-Combaret C. Ancient wheat varieties have a higher ability to interact with plant growth-promoting rhizobacteria. Plant Cell Environ. 2020 Jan;43(1):246–60.

71. Tadele Z. Orphan crops: their importance and the urgency of improvement. Planta. 2019 Sep;250(3):677–94.

72. Mabhaudhi T, Chimonyo VGP, Hlahla S, Massawe F, Mayes S, Nhamo L, et al. Prospects of orphan crops in climate change. Planta. 2019 Sep;250(3):695–708.

73. Stamp P, Visser R. The twenty-first century, the century of plant breeding. Euphytica. 2012 Aug;186(3):585–91.

74. Royer R. Doriane's Blog — Uncovering the Top Seed Companies [Internet]. 2024 [cited 2024 Jul 9]. Available from: https://www.doriane.com/blog/top-seed-companies

75. Quijano A. Coloniality of power and eurocentrism in latin america. International Sociology. 2000 Jun;15(2):215–32.

76. Eshel A, Beeckman T. Plant roots: the hidden half. books.google.com; 2013.

77. Enagbonma BJ, Fadiji AE, Ayangbenro AS, Babalola OO. Communication between Plants and Rhizosphere Microbiome: Exploring the Root Microbiome for Sustainable Agriculture. Microorganisms. 2023 Aug 3;11(8).

78. Tajima R. Importance of individual root traits to understand crop root system in agronomic and environmental contexts. Breed Sci. 2021 Feb;71(1):13–9.

79. Hochman G, Zilberman D. Corn ethanol and U.S. biofuel policy 10 years later: A quantitative assessment. Am J Agric Econ. 2018 Mar 1;100(2):570–84.

80. Robertson GP, Hamilton SK, Barham BL, Dale BE, Izaurralde RC, Jackson RD, et al. Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes. Science. 2017 Jun 30;356(6345).

81. Yang P, Cai X, Khanna M. Farmers' heterogeneous perceptions of marginal land for biofuel crops in US Midwestern states considering biophysical and socioeconomic factors. Glob Change Biol Bioenergy. 2021 May;13(5):849–61.

82. Bitterlich M, Franken P, Graefe J. Arbuscular mycorrhiza improves substrate hydraulic conductivity in the plant available moisture range under root growth exclusion. Front Plant Sci. 2018 Mar 7;9:301.

83. Boddey RM, Urquiaga S, Alves BJR, Reis V. Endophytic nitrogen fixation in sugarcane: present knowledge and future applications. Plant Soil. 2003 May;252(1):139–49.

84. Hartman K, Tringe SG. Interactions between plants and soil shaping the root microbiome under abiotic stress. Biochem J. 2019 Oct 15;476(19):2705–24.

85. Keymer DP, Kent AD. Contribution of nitrogen fixation to first year *Miscanthus* × *giganteus*. Glob Change Biol Bioenergy. 2014 Sep;6(5):577–86.

86. Smith SE, Read DJ. Mycorrhizal Symbiosis. 1st ed. Academic Press; 1996.

87. York LM, Carminati A, Mooney SJ, Ritz K, Bennett MJ. The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. J Exp Bot. 2016 Jun;67(12):3629–43.

88. Bakker MG, Manter DK, Sheflin AM, Weir TL, Vivanco JM. Harnessing the rhizosphere microbiome through plant breeding and agricultural management. Plant Soil. 2012 Nov;360(1–2):1–13.

89. Cotrufo MF, Ranalli MG, Haddix ML, Six J, Lugato E. Soil carbon storage informed by particulate and mineral-associated organic matter. Nat Geosci. 2019 Nov 18;

90. Frey SD. Mycorrhizal fungi as mediators of soil organic matter dynamics. Annu Rev Ecol Evol Syst. 2019 Nov 2;50(1):237–59.

91. Yang Y, Tilman D. Soil and root carbon storage is key to climate benefits of bioenergy crops. Biofuel Res J. 2020 Jun 1;7(2):1143–8.

92. De Deyn GB, Cornelissen JHC, Bardgett RD. Plant functional traits and soil carbon sequestration in contrasting biomes. Ecol Lett. 2008 May;11(5):516–31.

93. Jackson LE, Bowles TM, Hodson AK, Lazcano C. Soil microbial-root and microbial-rhizosphere processes to increase nitrogen availability and retention in agroecosystems. Current Opinion in Environmental Sustainability. 2012 Nov;4(5):517–22.

94. Simon PL, de Klein CAM, Worth W, Rutherford AJ, Dieckow J. The efficacy of Plantago lanceolata for mitigating nitrous oxide emissions from cattle urine patches. Sci Total Environ. 2019 Nov 15;691:430–41.

95. Coskun D, Britto DT, Shi W, Kronzucker HJ. How plant root exudates shape the nitrogen cycle. Trends Plant Sci. 2017 Aug;22(8):661–73.

96. Karwat H, Moreta D, Arango J, Núñez J, Rao I, Rincón Á, et al. Residual effect of BNI by Brachiaria humidicola pasture on nitrogen recovery and grain yield of subsequent maize. Plant Soil. 2017 Sep 12;1–18.

97. Villegas D, Arevalo A, Nuñez J, Mazabel J, Subbarao G, Rao I, et al. Biological Nitrification Inhibition (BNI): Phenotyping of a Core Germplasm Collection of the Tropical Forage Grass Megathyrsus maximus Under Greenhouse Conditions. Front Plant Sci. 2020 Jun 12;11:820.

98. Guyonnet JP, Vautrin F, Meiffren G, Labois C, Cantarel AAM, Michalet S, et al. The effects of plant nutritional strategy on soil microbial denitrification activity through rhizosphere primary metabolites. FEMS Microbiol Ecol. 2017 Apr 1;93(4).

99. Henry S, Texier S, Hallet S, Bru D, Dambreville C, Chèneby D, et al. Disentangling the rhizosphere effect on nitrate reducers and denitrifiers: insight into the role of root exudates. Environ Microbiol. 2008 Nov;10(11):3082–92.

100. Gomez-Casanovas N, DeLucia NJ, Hudiburg TW, Bernacchi CJ, DeLucia EH. Conversion of grazed pastures to energy cane as a biofuel feedstock alters the emission of GHGs from soils in Southeastern United States. Biomass and Bioenergy. 2018 Jan;108:312–22.

101. Bridgham SD, Cadillo-Quiroz H, Keller JK, Zhuang Q. Methane emissions from wetlands: biogeochemical, microbial, and modeling perspectives from local to global scales. Glob Chang Biol. 2013 May;19(5):1325–46.

102. Hodge A. Root decisions. Plant Cell Environ. 2009 Jun;32(6):628–40.

103. Germon A, Laclau J-P, Robin A, Jourdan C. Tamm Review: Deep fine roots in forest ecosystems: Why dig deeper? Forest Ecology and Management. 2020 Jun;466:118135.

104. Jackson RB, Canadell J, Ehleringer JR, Mooney HA, Sala OE, Schulze ED. A global analysis of root distributions for terrestrial biomes. Oecologia. 1996 Nov;108(3):389–411.

105. Pierret A, Maeght J-L, Clément C, Montoroi J-P, Hartmann C, Gonkhamdee S. Understanding deep roots and their functions in ecosystems: an advocacy for more unconventional research. Ann Bot. 2016 Oct 1;118(4):621–35.

106. Freycon V, Wonkam C, Fayolle A, Laclau J-P, Lucot E, Jourdan C, et al. Tree roots can penetrate deeply in African semi-deciduous rain forests: evidence from two common soil types. J Trop Ecol. 2015 Jan;31(1):13–23.

107. Zeng X, Dai Y-J, Dickinson RE, Shaikh M. The role of root distribution for climate simulation over land. Geophys Res Lett. 1998 Dec 15;25(24):4533–6.

108. Kleidon A, Heimann M. Assessing the role of deep rooted vegetation in the climate system with model simulations: mechanism, comparison to observations and implications for Amazonian deforestation. Clim Dyn. 2000 Feb 4;16(2–3):183–99.

109. York LM, Nord EA, Lynch JP. Integration of root phenes for soil resource acquisition. Front Plant Sci. 2013 Sep 12;4:355.

110. Teeri JA, Raven PH. A national ecological observatory network. Science. 2002 Dec 6;298(5600):1893–1893.

111. Schneider HM, Lynch JP. Should root plasticity be a crop breeding target? Front Plant Sci. 2020 May 15;11:546.

112. Freschet GT, Roumet C, Comas LH, Weemstra M, Bengough AG, Rewald B, et al. Root traits as drivers of plant and ecosystem functioning: current understanding, pitfalls and future research needs. New Phytol. 2021 Nov;232(3):1123–58.

113. Araus JL, Cairns JE. Field high-throughput phenotyping: the new crop breeding frontier. Trends Plant Sci. 2014 Jan;19(1):52–61.

114. Colombi T, Herrmann AM, Vallenback P, Keller T. Cortical cell diameter is key to energy costs of root growth in wheat. Plant Physiol. 2019 Aug;180(4):2049–60.

115. Guo H, Ayalew H, Seethepalli A, Dhakal K, Griffiths M, Ma X-F, et al. Functional phenomics and genetics of the root economics space in winter wheat using high-throughput phenotyping of respiration and architecture. New Phytol. 2021 Oct;232(1):98–112.

116. Comas LH, Becker SR, Cruz VMV, Byrne PF, Dierig DA. Root traits contributing to plant productivity under drought. Front Plant Sci. 2013 Nov 5;4:442.

117. Lynch JP. Root phenotypes for improved nutrient capture: an underexploited opportunity for global agriculture. New Phytol. 2019 Jul;223(2):548–64.

118. Guo H, York LM. Maize with fewer nodal roots allocates mass to more lateral and deep roots that improve nitrogen uptake and shoot growth. J Exp Bot. 2019 Oct 15;70(19):5299–309.

119. Klein SP, Schneider HM, Perkins AC, Brown KM, Lynch JP. Multiple Integrated Root Phenotypes Are Associated with Improved Drought Tolerance. Plant Physiol. 2020 Jul;183(3):1011–25.

120. Wu R, Grissom JE, McKeand SE, O'Malley DM. Phenotypic plasticity of fine root growth increases plant productivity in pine seedlings. BMC Ecol. 2004 Sep 7;4:14.

121. Ryan MG, Law BE. Interpreting, measuring, and modeling soil respiration. Biogeochemistry. 2005 Mar;73(1):3–27.

122. Kim D, Lee M-I, Seo E. Improvement of soil respiration parameterization in a dynamic global vegetation model and its impact on the simulation of terrestrial carbon fluxes. J Clim. 2019 Jan;32(1):127–43.

123. Kattge J, Bönisch G, Díaz S, Lavorel S, Prentice IC, Leadley P, et al. TRY plant trait database - enhanced coverage and open access. Glob Chang Biol. 2020 Jan;26(1):119–88.

124. Iversen CM, McCormack ML, Powell AS, Blackwood CB, Freschet GT, Kattge J, et al. A global Fine-Root Ecology Database to address below-ground challenges in plant ecology. New Phytol. 2017 Jul;215(1):15–26.

125. McKown AD, Klápště J, Guy RD, Geraldes A, Porth I, Hannemann J, et al. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. New Phytol. 2014 Jul;203(2):535–53.

126. Chhetri HB, Macaya-Sanz D, Kainer D, Biswal AK, Evans LM, Chen J-G, et al. Multitrait genome-wide association analysis of Populus trichocarpa identifies key polymorphisms controlling morphological and physiological traits. New Phytol. 2019 Jul;223(1):293–309.

127. Atkinson JA, Pound MP, Bennett MJ, Wells DM. Uncovering the hidden half of plants using new advances in root phenotyping. Curr Opin Biotechnol. 2019 Feb;55:1–8.

128. York LM. Functional phenomics: an emerging field integrating high-throughput phenotyping, physiology, and bioinformatics. J Exp Bot. 2019 Jan 7;70(2):379–86.

129. Lynch JP. Root architecture and plant productivity. Plant Physiol. 1995 Sep;109(1):7–13.

130. Zobel RW, Waisel Y. A plant root system architectural taxonomy: A framework for root nomenclature. Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology. 2010 Jun;144(2):507–12.

131. Pregitzer KS, DeForest JL, Burton AJ, Allen MF, Ruess RW, Hendrick RL. FINE ROOT ARCHITECTURE OF NINE NORTH AMERICAN TREES. Ecological Monographs. 2002 May 1;

132. Danjon F, Reubens B. Assessing and analyzing 3D architecture of woody root systems, a review of methods and applications in tree and soil stability, resource acquisition and allocation. Plant Soil. 2008 Feb;303(1–2):1–34.

133. Wahid PA. A system of classification of woody perennials based on their root activity patterns. Agroforestry Systems. 2000 Jul;

134. Cumming JR, Zawaski C, Desai S, Collart FR. Phosphorus disequilibrium in the tripartite plant-ectomycorrhiza-plant growth promoting rhizobacterial association. J Soil Sci Plant Nutr. 2015;15(ahead):0–0.

135. Bengough AG, Loades K, McKenzie BM. Root hairs aid soil penetration by anchoring the root surface to pore walls. J Exp Bot. 2016 Feb;67(4):1071–8.

136. Haling RE, Yang Z, Shadwell N, Culvenor RA, Stefanski A, Ryan MH, et al. Root morphological traits that determine phosphorus-acquisition efficiency and critical external phosphorus requirement in pasture species. Funct Plant Biol. 2016 Sep;43(9):815–26.

137. Carminati A, Passioura JB, Zarebanadkouki M, Ahmed MA, Ryan PR, Watt M, et al. Root hairs enable high transpiration rates in drying soils. New Phytol. 2017 Nov;216(3):771–81.

138. Ruiz S, Koebernick N, Duncan S, Fletcher DM, Scotson C, Boghi A, et al. Significance of root hairs at the field scale - modelling root water and phosphorus uptake under different field conditions. Plant Soil. 2020;447(1):281–304.

139. Cai G, Carminati A, Abdalla M, Ahmed MA. Soil textures rather than root hairs dominate water uptake and soil-plant hydraulics under drought. Plant Physiol. 2021 Oct 5;187(2):858–72.

140. Bates TR, Lynch JP. Plant growth and phosphorus accumulation of wild type and two root hair mutants of Arabidopsis thaliana (Brassicaceae). Am J Bot. 2000 Jul;87(7):958–63.

141. Haling RE, Brown LK, Bengough AG, Young IM, Hallett PD, White PJ, et al. Root hairs improve root penetration, root-soil contact, and phosphorus acquisition in soils of different strength. J Exp Bot. 2013 Sep;64(12):3711–21.

142. Ma JF, Goto S, Tamai K, Ichii M. Role of root hairs and lateral roots in silicon uptake by rice. Plant Physiol. 2001 Dec;127(4):1773–80.

143. McElrone AJ, Pockman WT, Martinez-Vilalta J, Jackson RB. Variation in xylem structure and function in stems and roots of trees to 20 m depth. New Phytol. 2004 Sep;163(3):507–17.

144. Kirfel K, Leuschner C, Hertel D, Schuldt B. Influence of Root Diameter and Soil Depth on the Xylem Anatomy of Fine- to Medium-Sized Roots of Mature Beech Trees in the Top- and Subsoil. Front Plant Sci. 2017 Jul 24;8:1194.

145. Tyree MT, Cochard H, Davis SD. Biophysical Perspectives of Xylem Evolution: is there a Tradeoff of Hydraulic Efficiency for Vulnerability to Dysfunction? IAWA. 1994 Jan 1;15(4):335–60.

146. Galindo-Castañeda T, Brown KM, Lynch JP. Reduced root cortical burden improves growth and grain yield under low phosphorus availability in maize. Plant Cell Environ. 2018 Jul;41(7):1579–92.

147. Griffiths M, York LM. Targeting root ion uptake kinetics to increase plant productivity and nutrient use efficiency. Plant Physiol. 2020 Apr;182(4):1854–68.

148. Griffiths M, Roy S, Guo H, Seethepalli A, Huhman D, Ge Y, et al. A multiple ion-uptake phenotyping platform reveals shared mechanisms affecting nutrient uptake by roots. Plant Physiol. 2021 Apr 2;185(3):781–95.

149. Watt M, McCully ME, Canny MJ. Formation and stabilization of Rhizosheaths of *Zea mays* L. (effect of soil water content). Plant Physiol. 1994 Sep;106(1):179–86.

150. Burak E, Quinton JN, Dodd IC. Root hairs are the most important root trait for rhizosheath formation of barley (Hordeum vulgare), maize (Zea mays) and Lotus japonicus (Gifu). Ann Bot. 2021 Jul 28;128(1):45–57.

151. Barré P, Hallett PD. Rheological stabilization of wet soils by model root and fungal exudates depends on clay mineralogy. Eur J Soil Sci. 2009 Aug;60(4):525–38.

152. Haling RE, Brown LK, Bengough AG, Valentine TA, White PJ, Young IM, et al. Root hair length and rhizosheath mass depend on soil porosity, strength and water content in barley genotypes. Planta. 2014 Mar;239(3):643–51.

153. Liu T-Y, Chen M-X, Zhang Y, Zhu F-Y, Liu Y-G, Tian Y, et al. Comparative metabolite profiling of two switchgrass ecotypes reveals differences in drought stress responses and rhizosheath weight. Planta. 2019 Oct;250(4):1355–69.

154. Fu HM, Meng FY, Molatudi RL, Zhang BG. Sorghum and switchgrass as biofuel feedstocks on marginal lands in northern china. Bioenerg Res. 2016 Jun;9(2):633–42.

155. Tang C, Li S, Li M, Xie GH. Bioethanol potential of energy sorghum grown on marginal and arable lands. Front Plant Sci. 2018 Apr 9;9:440.

156. Schetter A, Lin C-H, Zumpf C, Jang C, Hoffmann L, Rooney W, et al. Genotype-Environment-Management Interactions in Biomass Yield and Feedstock Composition of Photoperiod-Sensitive Energy Sorghum. Bioenerg Res. 2022 Jun;15(2):1017–32.

157. Frank AB, Berdahl JD, Hanson JD, Liebig MA, Johnson HA. Biomass and carbon partitioning in switchgrass. Crop Sci. 2004;44(4):1391.

158. Ma Z, Wood CW, Bransby DI. Impacts of soil management on root characteristics of switchgrass. Biomass and Bioenergy. 2000 Feb;18(2):105–12.

159. Graaff M-A de, Six J, Jastrow JD, Schadt CW, Wullschleger SD. Variation in root architecture among switchgrass cultivars impacts root decomposition rates. Soil Biol Biochem. 2013 Mar;58:198–206.

160. McLaughlin, Bouton, Bransby, Conger. Perspectives on New Crops and New Uses. Amer Soc for Horticultural Sc; 1999.

161. Shinde S, Cumming JR, Collart FR, Noirot PH, Larsen PE. Pseudomonas fluorescens Transportome Is Linked to Strain-Specific Plant Growth Promotion in Aspen Seedlings under Nutrient Stress. Front Plant Sci. 2017 Mar 21;8:348.

162. Shinde S, Zerbs S, Collart FR, Cumming JR, Noirot P, Larsen PE. Pseudomonas fluorescens increases mycorrhization and modulates expression of antifungal defense response genes in roots of aspen seedlings. BMC Plant Biol. 2019 Jan 3;19(1):4.

163. Douglas GB, McIvor IR, Lloyd-West CM. Early root development of field-grown poplar: effects of planting material and genotype. N Z J For Sci. 2016 Dec;46(1):1.

164. Phillips CJ, Marden M, Suzanne LM. Observations of root growth of young poplar and willow planting types. N Z J For Sci. 2014 Dec;44(1):15.

165. Mc Carthy R, Löf M, Gardiner ES. Early root development of poplars (*Populus* spp.) in relation to moist and saturated soil conditions. Scandinavian Journal of Forest Research. 2017 Jun 21;7(2):1–8.

166. Dash M, Yordanov YS, Georgieva T, Tschaplinski TJ, Yordanova E, Busov V. Poplar PtabZIP1-like enhances lateral root formation and biomass growth under drought stress. Plant J. 2017 Feb 10;89(4):692–705.

167. Dash M, Yordanov YS, Georgieva T, Wei H, Busov V. Gene network analysis of poplar root transcriptome in response to drought stress identifies a PtaJAZ3PtaRAP2.6-centered hierarchical network. PLoS ONE. 2018 Dec 12;13(12):e0208560.

168. Quinn LD, Straker KC, Guo J, Kim S, Thapa S, Kling G, et al. Stress-Tolerant Feedstocks for Sustainable Bioenergy Production on Marginal Land. Bioenerg Res. 2015 Sep;8(3):1081–100.

169. Yu F, Liang K, Han X, Du D, Pan Z, Qiu F. Major natural genetic variation contributes to waterlogging tolerance in maize seedlings. Mol Breeding. 2019 Jul;39(7):97.

170. Zhan A, Schneider H, Lynch JP. Reduced lateral root branching density improves drought tolerance in maize. Plant Physiol. 2015 Aug;168(4):1603–15.

171. Kameoka E, Suralta RR, Mitsuya S, Yamauchi A. Matching the Expression of Root Plasticity with Soil Moisture Availability Maximizes Production of Rice Plants Grown in an Experimental Sloping Bed having Soil Moisture Gradients. Plant Prod Sci. 2015 Jan;18(3):267–76.

172. Menge DM, Kameoka E, Kano-Nakata M, Yamauchi A, Asanuma S, Asai H, et al. Drought-induced root plasticity of two upland NERICA varieties under conditions with contrasting soil depth characteristics. Plant Prod Sci. 2016 Jul 2;19(3):389–400.

173. Yan X-L, Jia L, Dai T. Fine root morphology and growth in response to nitrogen addition through drip fertigation in a Populus × euramericana “Guariento” plantation over multiple years. Ann For Sci. 2019 Mar;76(1):13.

174. Domenicano S, Coll L, Messier C, Berninger F. Nitrogen forms affect root structure and water uptake in the hybrid poplar. New Forest. 2011 Nov;42(3):347–62.

175. Delhaize E, Rathjen TM, Cavanagh CR. The genetics of rhizosheath size in a multiparent mapping population of wheat. J Exp Bot. 2015 Aug;66(15):4527–36.

176. James RA, Weligama C, Verbyla K, Ryan PR, Rebetzke GJ, Rattey A, et al. Rhizosheaths on wheat grown in acid soils: phosphorus acquisition efficiency and genetic control. J Exp Bot. 2016 Jun;67(12):3709–18.

177. Lynch JP, Chimungu JG, Brown KM. Root anatomical phenes associated with water acquisition from drying soil: targets for crop improvement. J Exp Bot. 2014 Nov;65(21):6155–66.

178. Liu T-Y, Ye N, Song T, Cao Y, Gao B, Zhang D, et al. Rhizosheath formation and involvement in foxtail millet (Setaria italica) root growth under drought stress. J Integr Plant Biol. 2019 Apr;61(4):449–62.

179. Paez-Garcia A, Motes CM, Scheible W-R, Chen R, Blancaflor EB, Monteros MJ. Root traits and phenotyping strategies for plant improvement. Plants. 2015 Jun 15;4(2):334–55.

180. Topp CN, Bray AL, Ellis NA, Liu Z. How can we harness quantitative genetic variation in crop root systems for agricultural improvement? J Integr Plant Biol. 2016 Mar 11;58(3):213–25.

181. Andreasson F, Gonzalez M, Augusto L, Bakker MR. Comparison of ingrowth cores and ingrowth meshes in root studies: 3 years of data on Pinus pinaster and its understory. Trees. 2016 Apr;30(2):555–70.

182. Burridge JD, Black CK, Nord EA, Postma JA, Sidhu JS, York LM, et al. An Analysis of Soil Coring Strategies to Estimate Root Depth in Maize (Zea mays) and Common Bean (Phaseolus vulgaris). Plant Phenomics. 2020 Nov 8;2020:3252703.

183. Trachsel S, Kaeppler SM, Brown KM, Lynch JP. Shovelomics: high throughput phenotyping of maize (Zea mays L.) root architecture in the field. Plant Soil. 2011 Apr;341(1–2):75–87.

184. Seethepalli A, Guo H, Liu X, Griffiths M, Almtarfi H, Li Z, et al. Rhizovision crown: an integrated hardware and software platform for root crown phenotyping. Plant Phenomics. 2020 Feb 15;2020:3074916.

185. Möller B, Chen H, Schmidt T, Zieschank A, Patzak R, Türke M, et al. rhizoTrak: a flexible open source Fiji plugin for user-friendly manual annotation of time-series images from minirhizotrons. Plant Soil. 2019 Nov;444(1–2):519–34.

186. Smith AG, Han E, Petersen J, Olsen NAF, Giese C, Athmann M, et al. RootPainter: Deep Learning Segmentation of Biological Images with Corrective Annotation. BioRxiv. 2020 Apr 18;

187. Wang T, Rostamza M, Song Z, Wang L, McNickle G, Iyer-Pascuzzi AS, et al. SegRoot: A high throughput segmentation method for root image analysis. Computers and Electronics in Agriculture. 2019 Jul;162:845–54.

188. van Dusschoten D, Metzner R, Kochs J, Postma JA, Pflugfelder D, Bühler J, et al. Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging. Plant Physiol. 2016 Mar;170(3):1176–88.

189. Bagnall GC, Koonjoo N, Altobelli SA, Conradi MS, Fukushima E, Kuethe DO, et al. Low-field magnetic resonance imaging of roots in intact clayey and silty soils. Geoderma. 2020 Jul;370:114356.

190. Tracy SR, Roberts JA, Black CR, McNeill A, Davidson R, Mooney SJ. The X-factor: visualizing undisturbed root architecture in soils using X-ray computed tomography. J Exp Bot. 2010 Jan 4;61(2):311–3.

191. Guo L, Chen J, Cui X, Fan B, Lin H. Application of ground penetrating radar for coarse root detection and quantification: a review. Plant Soil. 2013 Jan;362(1–2):1–23.

192. Liu X, Dong X, Xue Q, Leskovar DI, Jifon J, Butnor JR, et al. Ground penetrating radar (GPR) detects fine roots of agricultural crops in the field. Plant Soil. 2018 Feb;423(1–2):517–31.

193. Peruzzo L, Chou C, Wu Y, Schmutz M, Mary B, Wagner FM, et al. Imaging of plant current pathways for non-invasive root Phenotyping using a newly developed electrical current source density approach. Plant Soil. 2020 May;450(1–2):567–84.

194. Armengaud P. EZ-Rhizo software. Plant Signal Behav. 2009 Feb;4(2):139–41.

195. Lobet G, Pagès L, Draye X. A novel image-analysis toolbox enabling quantitative analysis of root system architecture. Plant Physiol. 2011 Sep;157(1):29–39.

196. Pound MP, French AP, Atkinson JA, Wells DM, Bennett MJ, Pridmore T. RootNav: navigating images of complex root architectures. Plant Physiol. 2013 Aug;162(4):1802–14.

197. Pace J, Lee N, Naik HS, Ganapathysubramanian B, Lübberstedt T. Analysis of maize (Zea mays L.) seedling roots with the high-throughput image analysis tool ARIA (Automatic Root Image Analysis). PLoS ONE. 2014 Sep 24;9(9):e108255.

198. Rose L. Pitfalls in root trait calculations: how ignoring diameter heterogeneity can lead to overestimation of functional traits. Front Plant Sci. 2017 May 29;8:898.

199. Seethepalli A, Dhakal K, Griffiths M, Guo H, Freschet GT, York LM. RhizoVision Explorer: open-source software for root image analysis and measurement standardization. AoB Plants. 2021 Dec;13(6):plab056.

200. Silver WL, Miya RK. Global patterns in root decomposition: comparisons of climate and litter quality effects. Oecologia. 2001 Nov 1;129(3):407–19.

201. Austin EE, Castro HF, Sides KE, Schadt CW, Classen AT. Assessment of 10 years of CO2 fumigation on soil microbial communities and function in a sweetgum plantation. Soil Biol Biochem. 2009 Mar;41(3):514–20.

202. von Haden AC, Kucharik CJ, Jackson RD, Marín-Spiotta E. Litter quantity, litter chemistry, and soil texture control changes in soil organic carbon fractions under bioenergy cropping systems of the North Central U.S. Biogeochemistry. 2019 May;143(3):313–26.

203. Xu X, Sun Y, Sun J, Cao P, Wang Y, Chen HYH, et al. Cellulose dominantly affects soil fauna in the decomposition of forest litter: A meta-analysis. Geoderma. 2020 Nov;378:114620.

204. Poorter H, Niklas KJ, Reich PB, Oleksyn J, Poot P, Mommer L. Biomass allocation to leaves, stems and roots: meta-analyses of interspecific variation and environmental control. New Phytol. 2012 Jan;193(1):30–50.

205. Rasse DP, Rumpel C, Dignac M-F. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. Plant Soil. 2005 Feb;269(1–2):341–56.

206. Kong AYY, Six J. Tracing Root vs. Residue Carbon into Soils from Conventional and Alternative Cropping Systems. Soil Sci Soc Am J. 2010 Jul;74(4):1201–10.

207. Simpson AJ, Simpson MJ, Smith E, Kelleher BP. Microbially derived inputs to soil organic matter: are current estimates too low? Environ Sci Technol. 2007 Dec 1;41(23):8070–6.

208. Grandy AS, Neff JC. Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function. Sci Total Environ. 2008 Oct 15;404(2–3):297–307.

209. Kleber M, Johnson MG. Advances in understanding the molecular structure of soil organic matter. Advances in Agronomy v106. Elsevier; 2010. p. 77–142.

210. Schmidt MWI, Torn MS, Abiven S, Dittmar T, Guggenberger G, Janssens IA, et al. Persistence of soil organic matter as an ecosystem property. Nature. 2011 Oct 6;478(7367):49–56.

211. Farooq M, Wahid A, Kobayashi N, Fujita D, Basra SMA. Plant drought stress: effects, mechanisms and management. In: Lichtfouse E, Navarrete M, Debaeke P, Véronique S, Alberola C, editors. Sustainable Agriculture. Dordrecht: Springer Netherlands; 2009. p. 153–88.

212. Michaletti A, Naghavi MR, Toorchi M, Zolla L, Rinalducci S. Metabolomics and proteomics reveal drought-stress responses of leaf tissues from spring-wheat. Sci Rep. 2018 Apr 9;8(1):5710.

213. Bogeat-Triboulot M-B, Brosché M, Renaut J, Jouve L, Le Thiec D, Fayyaz P, et al. Gradual soil water depletion results in reversible changes of gene expression, protein profiles, ecophysiology, and growth performance in Populus euphratica, a poplar growing in arid regions. Plant Physiol. 2007 Feb;143(2):876–92.

214. Hamanishi ET, Raj S, Wilkins O, Thomas BR, Mansfield SD, Plant AL, et al. Intraspecific variation in the Populus balsamifera drought transcriptome. Plant Cell Environ. 2010 Oct;33(10):1742–55.

215. Hamanishi ET, Barchet GLH, Dauwe R, Mansfield SD, Campbell MM. Poplar trees reconfigure the transcriptome and metabolome in response to drought in a

genotype- and time-of-day-dependent manner. BMC Genomics. 2015 Apr 21;16(1):329.

216. Tschaplinski TJ, Abraham PE, Jawdy SS, Gunter LE, Martin MZ, Engle NL, et al. The nature of the progression of drought stress drives differential metabolomic responses in Populus deltoides. Ann Bot. 2019 Oct 29;124(4):617–26.

217. Ding N, Huertas R, Torres-Jerez I, Liu W, Watson B, Scheible W-R, et al. Transcriptional, metabolic, physiological and developmental responses of switchgrass to phosphorus limitation. Plant Cell Environ. 2021 Jan;44(1):186–202.

218. Regier N, Streb S, Cocozza C, Schaub M, Cherubini P, Zeeman SC, et al. Drought tolerance of two black poplar (Populus nigra L.) clones: contribution of carbohydrates and oxidative stress defence. Plant Cell Environ. 2009 Dec;32(12):1724–36.

219. Popović BM, Štajner D, Ždero-Pavlović R, Tumbas-Šaponjac V, Čanadanović-Brunet J, Orlović S. Water stress induces changes in polyphenol profile and antioxidant capacity in poplar plants (Populus spp.). Plant Physiol Biochem. 2016 Aug;105:242–50.

220. Clemensen AK, Provenza FD, Hendrickson JR, Grusak MA. Ecological Implications of Plant Secondary Metabolites - Phytochemical Diversity Can Enhance Agricultural Sustainability. Front Sustain Food Syst. 2020 Nov 19;4.

221. Uren. The Rhizosphere: Biochemistry and Organic Substances at the Soil-Plant Interface, Second Edition (Books in Soils, Plants, and the Environment). 2nd ed. Boca Raton, FL: CRC Press; 2007.

222. Nguyen C. Rhizodeposition of organic C by plants: mechanisms and controls. Agronomie. 2003 Jul;23(5–6):375–96.

223. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. The role of root exudates in rhizosphere interactions with plants and other organisms. Annu Rev Plant Biol. 2006;57:233–66.

224. Baudoin E, Benizri E, Guckert A. Impact of artificial root exudates on the bacterial community structure in bulk soil and maize rhizosphere. Soil Biol Biochem. 2003 Sep;35(9):1183–92.

225. Canarini A, Kaiser C, Merchant A, Richter A, Wanek W. Root exudation of primary metabolites: mechanisms and their roles in plant responses to environmental stimuli. Front Plant Sci. 2019 Feb 21;10:157.

226. Sasse J, Martinoia E, Northen T. Feed your friends: do plant exudates shape the root microbiome? Trends Plant Sci. 2018 Jan;23(1):25–41.

227. Shi S, Richardson AE, O'Callaghan M, DeAngelis KM, Jones EE, Stewart A, et al. Effects of selected root exudate components on soil bacterial communities. FEMS Microbiol Ecol. 2011 Sep;77(3):600–10.

228. Rolfe SA, Griffiths J, Ton J. Crying out for help with root exudates: adaptive mechanisms by which stressed plants assemble health-promoting soil microbiomes. Curr Opin Microbiol. 2019 Jun;49:73–82.

229. Canarini A, Mariotte P, Ingram L, Merchant A, Dijkstra FA. Mineral-Associated Soil Carbon is Resistant to Drought but Sensitive to Legumes and Microbial Biomass in an Australian Grassland. Ecosystems. 2018;21(2):349–59.

230. Dakora FD, Phillips DA. Root exudates as mediators of mineral acquisition in low-nutrient environments. In: Adu-Gyamfi JJ, editor. Food Security in Nutrient-Stressed Environments: Exploiting Plants' Genetic Capabilities. Dordrecht: Springer Netherlands; 2002. p. 201–13.

231. de Vries FT, Williams A, Stringer F, Willcocks R, McEwing R, Langridge H, et al. Changes in root-exudate-induced respiration reveal a novel mechanism through which drought affects ecosystem carbon cycling. New Phytol. 2019 Oct;224(1):132–45.

232. Preece C, Peñuelas J. Rhizodeposition under drought and consequences for soil communities and ecosystem resilience. Plant Soil. 2016 Dec;409(1–2):1–17.

233. Gargallo-Garriga A, Preece C, Sardans J, Oravec M, Urban O, Peñuelas J. Root exudate metabolomes change under drought and show limited capacity for recovery. Sci Rep. 2018 Aug 23;8(1):12696.

234. Carvalhais LC, Dennis PG, Fedoseyenko D, Hajirezaei M-R, Borriss R, von Wirén N. Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. Z Pflanzenernähr Bodenk. 2011 Feb;174(1):3–11.

235. Desai S, Naik D, Cumming JR. The influence of phosphorus availability and Laccaria bicolor symbiosis on phosphate acquisition, antioxidant enzyme activity, and rhizospheric carbon flux in Populus tremuloides. Mycorrhiza. 2014 Jul;24(5):369–82.

236. Higa A, Mori Y, Kitamura Y. Iron deficiency induces changes in riboflavin secretion and the mitochondrial electron transport chain in hairy roots of Hyoscyamus albus. J Plant Physiol. 2010 Jul 15;167(11):870–8.

237. Kraffczyk I, Trolldenier G, Beringer H. Soluble root exudates of maize: Influence of potassium supply and rhizosphere microorganisms. Soil Biol Biochem. 1984 Jan;16(4):315–22.

238. Naik D, Smith E, Cumming JR. Rhizosphere carbon deposition, oxidative stress and nutritional changes in two poplar species exposed to aluminum. Tree Physiol. 2009 Mar;29(3):423–36.

239. Vranova V, Rejsek K, Skene KR, Janous D, Formanek P. Methods of collection of plant root exudates in relation to plant metabolism and purpose: A review. Z Pflanzenernähr Bodenk. 2013 Apr;176(2):175–99.

240. Gschwendtner S, Esperschütz J, Buegger F, Reichmann M, Müller M, Munch JC, et al. Effects of genetically modified starch metabolism in potato plants on photosynthate fluxes into the rhizosphere and on microbial degraders of root exudates. FEMS Microbiol Ecol. 2011 Jun;76(3):564–75.

241. Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, et al. Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. Proc Natl Acad Sci USA. 2020 Feb 18;117(7):3874–83.

242. Klugh KR, Cumming JR. Variations in organic acid exudation and aluminum resistance among arbuscular mycorrhizal species colonizing Liriodendron tulipifera. Tree Physiol. 2007 Aug;27(8):1103–12.

243. Klugh-Stewart K, Cumming JR. Organic acid exudation by mycorrhizal Andropogon virginicus L. (broomsedge) roots in response to aluminum. Soil Biol Biochem. 2009 Feb;41(2):367–73.

244. Eilers KG, Lauber CL, Knight R, Fierer N. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. Soil Biol Biochem. 2010 Jun;42(6):896–903.

245. Haichar F el Z, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J, et al. Plant host habitat and root exudates shape soil bacterial community structure. ISME J. 2008 Dec;2(12):1221–30.

246. Kuzyakov Y, Friedel JK, Stahr K. Review of mechanisms and quantification of priming effects. Soil Biol Biochem. 2000 Oct;32(11–12):1485–98.

247. Bradford MA, Fierer N, Reynolds JF. Soil carbon stocks in experimental mesocosms are dependent on the rate of labile carbon, nitrogen and phosphorus inputs to soils. Funct Ecol. 2008 Dec;22(6):964–74.

248. Cheng W, Coleman DC. Effect of living roots on soil organic matter decomposition. Soil Biol Biochem. 1990 Jan;22(6):781–7.

249. Cheng W, Kuzyakov Y. Root effects on soil organic matter decomposition. In: Zobel RW, Wright SF, editors. Roots and Soil Management: Interactions between Roots and the Soil. Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America; 2005. p. 119–43.

250. De Nobili M, Contin M, Mondini C, Brookes PC. Soil microbial biomass is triggered into activity by trace amounts of substrate. Soil Biol Biochem. 2001 Jul;33(9):1163–70.

251. BODDY E, HILL P, FARRAR J, JONES D. Fast turnover of low molecular weight components of the dissolved organic carbon pool of temperate grassland field soils. Soil Biol Biochem. 2007 Apr;39(4):827–35.

252. Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol. 2007 Nov;73(21):7059–66.

253. Wieder WR, Boehnert J, Bonan GB. Evaluating soil biogeochemistry parameterizations in Earth system models with observations. Global Biogeochem Cycles. 2014 Mar;28(3):211–22.

254. Egenolf K, Verma S, Schöne J, Klaiber I, Arango J, Cadisch G, et al. Rhizosphere pH and cation-anion balance determine the exudation of nitrification inhibitor 3-epi-brachialactone suggesting release via secondary transport. Physiol Plant. 2021 May;172(1):116–23.

255. Elle O, Richter R, Vohland M, Weigelt A. Fine root lignin content is well predictable with near-infrared spectroscopy. Sci Rep. 2019 Apr 23;9(1):6396.

256. Clemente JS, Simpson MJ, Simpson AJ, Yanni SF, Whalen JK. Comparison of soil organic matter composition after incubation with maize leaves, roots, and stems. Geoderma. 2013 Jan;192:86–96.

257. Rencoret J, Marques G, Serrano O, Kaal J, Martínez AT, del Río JC, et al. Deciphering the Unique Structure and Acylation Pattern of *Posidonia oceanica* Lignin. ACS Sustain Chem Eng. 2020 Aug 24;8(33):12521–33.

258. Legner N, Meinen C, Rauber R. Root differentiation of agricultural plant cultivars and proveniences using FTIR spectroscopy. Front Plant Sci. 2018 Jun 5;9:748.

259. Li H, Bölscher T, Winnick M, Tfaily MM, Cardon ZG, Keiluweit M. Simple Plant and Microbial Exudates Destabilize Mineral-Associated Organic Matter via Multiple Pathways. Environ Sci Technol. 2021 Mar 2;55(5):3389–98.

260. Sasse J, Kosina SM, de Raad M, Jordan JS, Whiting K, Zhalnina K, et al. Root morphology and exudate availability are shaped by particle size and chemistry in *Brachypodium distachyon*. Plant Direct. 2020 Jul 2;4(7):e00207.

261. Sasse J, Kant J, Cole BJ, Klein AP, Arsova B, Schlaepfer P, et al. Multilab EcoFAB study shows highly reproducible physiology and depletion of soil metabolites by a model grass. New Phytol. 2019 Apr;222(2):1149–60.

262. Bakker PAHM, Berendsen RL, Doornbos RF, Wintermans PCA, Pieterse CMJ. The rhizosphere revisited: root microbiomics. Front Plant Sci. 2013 May 30;4:165.

263. Inderjit, Weston LA. Root exudates: an overview. In: de Kroon H, Visser EJW, editors. Root Ecology. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 235–55.

264. Li Y, Li Q, Guan G, Chen S. Phosphate solubilizing bacteria stimulate wheat rhizosphere and endosphere biological nitrogen fixation by improving phosphorus content. PeerJ. 2020 May 4;8:e9062.

265. Martin-Rivilla H, Garcia-Villaraco A, Ramos-Solano B, Gutierrez-Mañero FJ, Lucas JA. Bioeffectors as biotechnological tools to boost plant innate immunity: signal transduction pathways involved. Plants. 2020 Dec 8;9(12).

266. Wang F, Feng G. Arbuscular mycorrhizal fungi interactions in the rhizosphere. In: Gupta VVSR, Sharma AK, editors. Rhizosphere biology: interactions between microbes and plants. Singapore: Springer Singapore; 2021. p. 217–35.

267. Jiang F, Zhang L, Zhou J, George TS, Feng G. Arbuscular mycorrhizal fungi enhance mineralisation of organic phosphorus by carrying bacteria along their extraradical hyphae. New Phytol. 2021 Apr;230(1):304–15.

268. Asanuma S, Tanaka H, Yatazawa M. Rhizoplane microorganisms of rice seedlings as examined by scanning electron microscopy. Soil Sci Plant Nutr. 1979 Dec;25(4):539–51.

269. Balasundararajan V, Dananjeyan B. Occurrence of diversified N-acyl homoserine lactone mediated biofilm-forming bacteria in rice rhizoplane. J Basic Microbiol. 2019 Oct;59(10):1031–9.

270. Kumar A, Singh J. Biofilms forming microbes: diversity and potential application in plant–microbe interaction and plant growth. In: Yadav AN, Singh J, Rastegari AA, Yadav N, editors. Plant microbiomes for sustainable agriculture. Cham: Springer International Publishing; 2020. p. 173–97.

271. Sattelmacher B. The apoplast and its significance for plant mineral nutrition. New Phytol. 2001 Feb;149(2):167–92.

272. Provorov NA. [Plant-microbe symbioses as an evolutionary continuum]. Zh Obshch Biol. 2009;70(1):10–34.

273. Dini-Andreote F. Endophytes: the second layer of plant defense. Trends Plant Sci. 2020 Apr;25(4):319–22.

274. Berthelot C, Chalot M, Leyval C, Blaudez D. From darkness to light: emergence of the mysterious dark septate endophytes in plant growth promotion and stress alleviation. In: Hodkinson TR, Doohan FM, Saunders MJ, Murphy BR, editors. Endophytes for a growing world. Cambridge University Press; 2019. p. 143–64.

275. Doty SL, Oakley B, Xin G, Kang JW, Singleton G, Khan Z, et al. Diazotrophic endophytes of native black cottonwood and willow. Symbiosis. 2009 Feb;47(1):23–33.

276. Ghimire SR, Charlton ND, Craven KD. The Mycorrhizal Fungus, Sebacina vermifera, Enhances Seed Germination and Biomass Production in Switchgrass (Panicum virgatum L). Bioenerg Res. 2009 Jun;2(1–2):51–8.

277. Wang B, Seiler JR, Mei C. A microbial endophyte enhanced growth of switchgrass under two drought cycles improving leaf level physiology and leaf development. Environ Exp Bot. 2016 Feb;122:100–8.

278. Torres MS, White JF, Zhang X, Hinton DM, Bacon CW. Endophyte-mediated adjustments in host morphology and physiology and effects on host fitness traits in grasses. Fungal Ecol. 2012 Jun;5(3):322–30.

279. Babu AG, Shim J, Bang K-S, Shea PJ, Oh B-T. Trichoderma virens PDR-28: a heavy metal-tolerant and plant growth-promoting fungus for remediation and bioenergy crop production on mine tailing soil. J Environ Manage. 2014 Jan;132:129–34.

280. Plett JM, Daguerre Y, Wittulsky S, Vayssières A, Deveau A, Melton SJ, et al. Effector MiSSP7 of the mutualistic fungus Laccaria bicolor stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. Proc Natl Acad Sci USA. 2014 Jun 3;111(22):8299–304.

281. Rush TA, Puech-Pagès V, Bascaules A, Jargeat P, Maillet F, Haouy A, et al. Lipo-chitooligosaccharides as regulatory signals of fungal growth and development. Nat Commun. 2020 Aug 4;11(1):3897.

282. Kell DB. Breeding crop plants with deep roots: their role in sustainable carbon, nutrient and water sequestration. Ann Bot. 2011 Sep;108(3):407–18.

283. Moore CE, Haden AC, Burnham MB, Kantola IB, Gibson CD, Blakely BJ, et al. Ecosystem-scale biogeochemical fluxes from three bioenergy crop candidates: How energy sorghum compares to maize and miscanthus. Glob Change Biol Bioenergy. 2021 Mar;13(3):445–58.

284. Cruz C, Green JJ, Watson CA, Wilson F, Martins-Loução MA. Functional aspects of root architecture and mycorrhizal inoculation with respect to nutrient uptake capacity. Mycorrhiza. 2004 Jul;14(3):177–84.

285. Pervaiz ZH, Contreras J, Hupp BM, Lindenberger JH, Chen D, Zhang Q, et al. Root microbiome changes with root branching order and root chemistry in peach rhizosphere soil. Rhizosphere. 2020 Dec;16:100249.

286. Saleem M, Law AD, Sahib MR, Pervaiz ZH, Zhang Q. Impact of root system architecture on rhizosphere and root microbiome. Rhizosphere. 2018 Jun;6:47–51.

287. Lawson CE, Harcombe WR, Hatzenpichler R, Lindemann SR, Löffler FE, O'Malley MA, et al. Common principles and best practices for engineering microbiomes. Nat Rev Microbiol. 2019 Dec;17(12):725–41.

288. Compant S, Clément C, Sessitsch A. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem. 2010 May;42(5):669–78.

289. Labbé J, Muchero W, Czarnecki O, Wang J, Wang X, Bryan AC, et al. Mediation of plant-mycorrhizal interaction by a lectin receptor-like kinase. Nat Plants. 2019 Jul 8;5(7):676–80.

290. Sun Y, Qiao Z, Muchero W, Chen J-G. Lectin Receptor-Like Kinases: The Sensor and Mediator at the Plant Cell Surface. Front Plant Sci. 2020 Dec 10;11:596301.

291. R Cope K, B Irving T, Chakraborty S, Ané J-M. Perception of lipo-chitooligosaccharides by the bioenergy crop Populus. Plant Signal Behav. 2021 Jun 3;16(6):1903758.

292. Beckers B, Op De Beeck M, Thijs S, Truyens S, Weyens N, Boerjan W, et al. Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. Front Microbiol. 2016 May 13;7:650.

293. Gutjahr C, Sawers RJH, Marti G, Andrés-Hernández L, Yang S-Y, Casieri L, et al. Transcriptome diversity among rice root types during asymbiosis and interaction with arbuscular mycorrhizal fungi. Proc Natl Acad Sci USA. 2015 May 26;112(21):6754–9.

294. Salas-González I, Reyt G, Flis P, Custódio V, Gopaulchan D, Bakhoum N, et al. Coordination between microbiota and root endodermis supports plant mineral nutrient homeostasis. Science. 2021 Jan 8;371(6525).

295. Lin J, Roswanjaya YP, Kohlen W, Stougaard J, Reid D. Nitrate inhibits nodule organogenesis through inhibition of cytokinin biosynthesis in *Lotus japonicus*. BioRxiv. 2020 Nov 4;

296. Lin C-Y, Eudes A. Strategies for the production of biochemicals in bioenergy crops. Biotechnol Biofuels. 2020 Apr 15;13:71.

297. Lin C-Y, Vuu KM, Amer B, Shih PM, Baidoo EEK, Scheller HV, et al. In-planta production of the biodegradable polyester precursor 2-pyrone-4,6-dicarboxylic acid (PDC): Stacking reduced biomass recalcitrance with value-added co-product. Metab Eng. 2021 Jul;66:148–56.

298. Bailey J, Cass J, Gasper J, Ngo N-D, Wiggins P, Manoil C. Essential gene deletions producing gigantic bacteria. PLoS Genet. 2019 Jun 10;15(6):e1008195.

299. Dynarski KA, Bossio DA, Scow KM. Dynamic stability of soil carbon: reassessing the “permanence” of soil carbon sequestration. Front Environ Sci. 2020 Nov 13;8.

300. Weigelt A, Mommer L, Andraczek K, Iversen CM, Bergmann J, Bruelheide H, et al. An integrated framework of plant form and function: the belowground perspective. New Phytol. 2021 Oct;232(1):42–59.

301. O’Toole JC, Bland WL. Genotypic variation in crop plant root systems. Advances in agronomy volume 41. Elsevier; 1987. p. 91–145.

302. Brunner I, Herzog C, Galiano L, Gessler A. Plasticity of Fine-Root Traits Under Long-Term Irrigation of a Water-Limited Scots Pine Forest. Front Plant Sci. 2019 Jun 4;10:701.

303. Downie HF, Adu MO, Schmidt S, Otten W, Dupuy LX, White PJ, et al. Challenges and opportunities for quantifying roots and rhizosphere interactions through imaging and image analysis. Plant Cell Environ. 2015 Jul;38(7):1213–32.

304. Hodge A. The plastic plant: root responses to heterogeneous supplies of nutrients. New Phytol. 2004 Apr;162(1):9–24.

305. Tracy SR, Nagel KA, Postma JA, Fassbender H, Wasson A, Watt M. Crop Improvement from Phenotyping Roots: Highlights Reveal Expanding Opportunities. Trends Plant Sci. 2020 Jan;25(1):105–18.

306. Lawton. The Ecotron: a controlled environmental facility for the investigation of population and ecosystem processes. Phil Trans R Soc Lond B. 1993 Jul 29;341(1296):181–94.

307. Singer E, Vogel JP, Northen T, Mungall CJ, Juenger TE. Novel and Emerging Capabilities that Can Provide a Holistic Understanding of the Plant Root Microbiome. Phytobiomes Journal. 2021 Jan 18;PBIOMES-05-20-0.

308. Dowd T, McInturf S, Li M, Topp CN. Rated-M for mesocosm: allowing the multimodal analysis of mature root systems in 3D. Emerg Top Life Sci. 2021 May 21;5(2):249–60.

309. Zengler K, Hofmockel K, Baliga NS, Behie SW, Bernstein HC, Brown JB, et al. EcoFABs: advancing microbiome science through standardized fabricated ecosystems. Nat Methods. 2019 Jul;16(7):567–71.

310. Paliy O, Shankar V. Application of multivariate statistical techniques in microbial ecology. Mol Ecol. 2016 Mar;25(5):1032–57.

311. Azodi CB, Tang J, Shiu S-H. Opening the black box: interpretable machine learning for geneticists. Trends Genet. 2020 Jun;36(6):442–55.

312. Gray SB, Strellner RS, Puthuval KK, Ng C, Shulman RE, Siebers MH, et al. Minirhizotron imaging reveals that nodulation of field-grown soybean is enhanced by free-air CO2 enrichment only when combined with drought stress. Funct Plant Biol. 2013 Mar;40(2):137–47.

313. York LM, Lynch JP. Intensive field phenotyping of maize (Zea mays L.) root crowns identifies phenes and phene integration associated with plant growth and nitrogen acquisition. J Exp Bot. 2015 Sep;66(18):5493–505.

314. Lobet G, Draye X, Périlleux C. An online database for plant image analysis software tools. Plant Methods. 2013 Oct 9;9(1):38.

315. Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Inf Serv Use. 2017 Mar 7;37(1):49–56.

316. Jansson JK, Hofmockel KS. The soil microbiome-from metagenomics to metaphenomics. Curr Opin Microbiol. 2018 Jun;43:162–8.

317. Intergovernmental Panel on Climate Change. Climate Change 2021: The physical science basis [Internet]. 2021 [cited 2024 Jul 11]. Available from: https://cir.nii.ac.jp/crid/1371413280484207233

318. Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, et al. Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci USA. 2013 Apr 16;110(16):6548–53.

319. Schlaeppi K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. Proc Natl Acad Sci USA. 2014 Jan 14;111(2):585–92.

320. Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, et al. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci USA. 2015 Feb 24;112(8):E911-20.

321. Naylor D, DeGraaf S, Purdom E, Coleman-Derr D. Drought and host selection influence bacterial community dynamics in the grass root microbiome. ISME J. 2017 Dec;11(12):2691–704.

322. Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, et al. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. Proc Natl Acad Sci USA. 2018 Jul 10;115(28):7368–73.

323. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, et al. PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science. 2015 Aug 21;349(6250):860–4.

324. Castrillo G, Teixeira PJPL, Paredes SH, Law TF, de Lorenzo L, Feltcher ME, et al. Root microbiota drive direct integration of phosphate stress and immunity. Nature. 2017 Mar 23;543(7646):513–8.

325. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, et al. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat Microbiol. 2018 Apr;3(4):470–80.

326. Brachi B, Morris GP, Borevitz JO. Genome-wide association studies in plants: the missing heritability is in the field. Genome Biol. 2011 Oct 28;12(10):232.

327. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 2010 Jun 3;465(7298):627–31.

328. Wu S, Tohge T, Cuadros-Inostroza Á, Tong H, Tenenboim H, Kooke R, et al. Mapping the arabidopsis metabolic landscape by untargeted metabolomics at different environmental conditions. Mol Plant. 2018 Jan 8;11(1):118–34.

329. Schaefer RJ, Michno J-M, Jeffers J, Hoekenga O, Dilkes B, Baxter I, et al. Integrating Coexpression Networks with GWAS to Prioritize Causal Genes in Maize. Plant Cell. 2018 Dec;30(12):2922–42.

330. Davenport ER, Cusanovich DA, Michelini K, Barreiro LB, Ober C, Gilad Y. Genome-Wide Association Studies of the Human Gut Microbiota. PLoS ONE. 2015 Nov 3;10(11):e0140301.

331. Wang J, Thingholm LB, Skiecevičienė J, Rausch P, Kummen M, Hov JR, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat Genet. 2016 Nov;48(11):1396–406.

332. Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, et al. Genome-wide association study of Arabidopsis thaliana leaf microbial community. Nat Commun. 2014 Nov 10;5:5320.

333. Wallace JG, Kremling KA, Kovar LL, Buckler ES. Quantitative genetics of the maize leaf microbiome. Phytobiomes Journal. 2018 Jan;2(4):208–24.

334. Roman-reyna V, Pinili D, Borjaa FN, Quibod IL, Groen SC, Mulyaningsih ES, et al. The rice leaf microbiome has a conserved community structure controlled by complex host-microbe interactions. BioRxiv. 2019 Apr 22;

335. Bergelson J, Mittelstrass J, Horton MW. Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome. Sci Rep. 2019 Jan 10;9(1):24.

336. Baerson SR, Dayan FE, Rimando AM, Nanayakkara NPD, Liu C-J, Schröder J, et al. A functional genomics investigation of allelochemical biosynthesis in Sorghum bicolor root hairs. J Biol Chem. 2008 Feb 8;283(6):3231–47.

337. Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, et al. Community Resources and Strategies for Association Mapping in Sorghum. Crop Sci. 2008;48(1):30.

338. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci USA. 2013 Jan 8;110(2):453–8.

339. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014 Jul;31(7):1929–36.

340. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007 Oct 1;23(19):2633–5.

341. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256–9.

342. Simmons T, Caddell DF, Deng S, Coleman-Derr D. Exploring the Root Microbiome: Extracting Bacterial Community Data from the Soil, Rhizosphere, and Root Endosphere. J Vis Exp. 2018 May 2;(135).

343. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019 Aug;37(8):852–7.

344. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011 Nov 1;27(21):2957–63.

345. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013 Oct;10(10):996–8.

346. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007 Aug;73(16):5261–7.

347. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE. 2011 Dec 14;6(12):e27310.

348. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013 Dec;10(12):1200–2.

349. Holland JB, Nyquist WE, Cervantes-Martínez CT. Estimating and Interpreting Heritability for Plant Breeding: An Update. In: Janick J, editor. Plant Breeding Reviews. Oxford, UK: John Wiley & Sons, Inc.; 2002. p. 9–112.

350. Lee SH, van der Werf JHJ. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics. 2016 May 1;32(9):1420–2.

351. Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat Stat. 2018 Mar;23:52–71.

352. Covarrubias-Pazaran G. Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS ONE. 2016 Jun 6;11(6):e0156744.

353. García-López R, Cornejo-Granados F, Lopez-Zavala AA, Sánchez-López F, Cota-Huízar A, Sotelo-Mundo RR, et al. Doing More with Less: A Comparison of 16S Hypervariable Regions in Search of Defining the Shrimp Microbiota. Microorganisms. 2020 Jan 17;8(1).

354. Kido T, Sikora-Wohlfeld W, Kawashima M, Kikuchi S, Kamatani N, Patwardhan A, et al. Are minor alleles more likely to be risk alleles? BMC Med Genomics. 2018 Jan 19;11(1):3.

355. Yu J, Holland JB, McMullen MD, Buckler ES. Genetic design and statistical power of nested association mapping in maize. Genetics. 2008 Jan;178(1):539–51.

356. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012 Jun 17;44(7):821–4.

357. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB. Vegan: community ecology package software. 2016.

358. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004 Jan 22;20(2):289–90.

359. Roberts DW, Roberts MDW. Package 'labdsv'. In: Ordination and multivariate. 2016.

360. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012 Jan;40(D1):D1178–86.

361. Xu L, Naylor D, Dong Z, Simmons T, Pierroz G, Hixson KK, et al. Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. Proc Natl Acad Sci USA. 2018 May 1;115(18):E4284–93.

362. Oberholster T, Vikram S, Cowan D, Valverde A. Key microbial taxa in the rhizosphere of sorghum and sunflower grown in crop rotation. Sci Total Environ. 2018 May 15;624:530–9.

363. Beilsmith K, Thoen MPM, Brachi B, Gloss AD, Khan MH, Bergelson J. Genome-wide association studies on the phyllosphere microbiome: Embracing complexity in host-microbe interactions. Plant J. 2019 Jan 5;97(1):164–81.

364. Swigonova Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, et al. On the tetraploid origin of the maize genome. Comp Funct Genomics. 2004;5(3):281–4.

365. Perneger TV. What's wrong with Bonferroni adjustments. BMJ. 1998 Apr 18;316(7139):1236–8.

366. Kaler AS, Purcell LC. Estimation of a significance threshold for genome-wide association studies. BMC Genomics. 2019 Jul 29;20(1):618.

367. Arouisse B, Korte A, van Eeuwijk F, Kruijer W. Imputation of 3 million SNPs in the Arabidopsis regional mapping population. Plant J. 2020 May;102(4):872–82.

368. Bodenhausen N, Horton MW, Bergelson J. Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana. PLoS ONE. 2013 Feb 15;8(2):e56329.

369. Copeland JK, Yuan L, Layeghifard M, Wang PW, Guttman DS. Seasonal community succession of the phyllosphere microbiome. Mol Plant Microbe Interact. 2015 Mar;28(3):274–85.

370. Badri DV, Chaparro JM, Zhang R, Shen Q, Vivanco JM. Application of natural blends of phytochemicals derived from the root exudates of Arabidopsis to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. J Biol Chem. 2013 Feb 15;288(7):4502–12.

371. Zhang N, Wang D, Liu Y, Li S, Shen Q, Zhang R. Effects of different plant root exudates and their organic acid components on chemotaxis, biofilm formation and colonization by beneficial rhizosphere-associated bacterial strains. Plant Soil. 2014 Jan;374(1–2):689–700.

372. Chen T, Nomura K, Wang X, Sohrabi R, Xu J, Yao L, et al. A plant genetic network for preventing dysbiosis in the phyllosphere. Nature. 2020 Apr 8;580(7805):653–7.

373. Angus AA, Agapakis CM, Fong S, Yerrapragada S, Estrada-de los Santos P, Yang P, et al. Plant-associated symbiotic Burkholderia species lack hallmark strategies required in mammalian pathogenesis. PLoS ONE. 2014 Jan 8;9(1):e83779.

374. Kim JK, Lee BL. Symbiotic factors in Burkholderia essential for establishing an association with the bean bug, Riptortus pedestris. Arch Insect Biochem Physiol. 2015 Jan;88(1):4–17.

375. Shu L, Brock DA, Geist KS, Miller JW, Queller DC, Strassmann JE, et al. Symbiont location, host fitness, and possible coadaptation in a symbiosis between social amoebae and bacteria. eLife. 2018 Dec 31;7.

376. Mannaa M, Park I, Seo Y-S. Genomic Features and Insights into the Taxonomy, Virulence, and Benevolence of Plant-Associated Burkholderia Species. Int J Mol Sci. 2018 Dec 29;20(1).

377. Erbilgin O, McDonald KL, Kerfeld CA. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic degradation of plant saccharides. Appl Environ Microbiol. 2014 Apr;80(7):2193–205.

378. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature. 2012 Aug 2;488(7409):91–5.

379. Pascale A, Proietti S, Pantelides IS, Stringlis IA. Modulation of the root microbiome by plant molecules: the basis for targeted disease suppression and plant growth promotion. Front Plant Sci. 2019;10:1741.

380. Khan GA, Vogiatzaki E, Glauser G, Poirier Y. Phosphate deficiency induces the jasmonate pathway and enhances resistance to insect herbivory. Plant Physiol. 2016 May;171(1):632–44.

381. Hiruma K, Gerlach N, Sacristán S, Nakano RT, Hacquard S, Kracher B, et al. Root Endophyte Colletotrichum tofieldiae Confers Plant Fitness Benefits that Are Phosphate Status Dependent. Cell. 2016 Apr 7;165(2):464–74.

382. Yamada K, Saijo Y, Nakagami H, Takano Y. Regulation of sugar transporter activity for antibacterial defense in Arabidopsis. Science. 2016 Dec 16;354(6318):1427–30.

383. Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nat Commun. 2016 Jul 12;7:12151.

384. Edwards JA, Santos-Medellín CM, Liechty ZS, Nguyen B, Lurie E, Eason S, et al. Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. PLoS Biol. 2018 Feb 23;16(2):e2003862.

385. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core Arabidopsis thaliana root microbiome. Nature. 2012 Aug 2;488(7409):86–90.

386. Haney CH, Samuel BS, Bush J, Ausubel FM. Associations with rhizosphere bacteria can confer an adaptive advantage to plants. Nat Plants. 2015 May 11;1(6).

387. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. Assembly and ecological function of the root microbiome across angiosperm plant species. Proc Natl Acad Sci USA. 2018 Feb 6;115(6):E1157–65.

388. Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna CV, Sánchez-León S, et al. High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. Plant J. 2017 Mar;89(6):1251–62.

389. Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, Bocobza S, Czosnek H, Aharoni A, et al. Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. Plant J. 2018 Jul;95(1):5–16.

390. Brückner K, Schäfer P, Weber E, Grützner R, Marillonnet S, Tissier A. A library of synthetic transcription activator-like effector-activated promoters for coordinated orthogonal gene expression in plants. Plant J. 2015 May;82(4):707–16.

391. Brophy JAN, Magallon KJ, Duan L, Zhong V, Ramachandran P, Kniazev K, et al. Synthetic genetic circuits as a means of reprogramming plant roots. Science. 2022 Aug 12;377(6607):747–51.

392. Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, et al. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc Natl Acad Sci USA. 2013 Aug 20;110(34):14024–9.

393. Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai Q-A, et al. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods. 2013 Apr;10(4):354–60.

394. Schultz BJ, Kim SY, Lau W, Sattely ES. Total Biosynthesis for Milligram-Scale Production of Etoposide Intermediates in a Plant Chassis. J Am Chem Soc. 2019 Dec 11;141(49):19231–5.

395. Marsian J, Lomonossoff GP. Molecular pharming - VLPs made in plants. Curr Opin Biotechnol. 2016 Feb;37:201–6.

396. Reed J, Stephenson MJ, Miettinen K, Brouwer B, Leveau A, Brett P, et al. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. Metab Eng. 2017 Jul 4;42:185–93.

397. Lloyd JPB, Ly F, Gong P, Pflueger J, Swain T, Pflueger C, et al. Synthetic memory circuits for stable cell reprogramming in plants. Nat Biotechnol. 2022 Dec;40(12):1862–72.

398. Belcher MS, Vuu KM, Zhou A, Mansoori N, Agosto Ramos A, Thompson MG, et al. Design of orthogonal regulatory systems for modulating gene expression in plants. Nat Chem Biol. 2020 Aug;16(8):857–65.

399. Schaumberg KA, Antunes MS, Kassaw TK, Xu W, Zalewski CS, Medford JI, et al. Quantitative characterization of genetic parts and circuits for plant synthetic biology. Nat Methods. 2016 Jan;13(1):94–100.

400. Jameel A, Noman M, Liu W, Ahmad N, Wang F, Li X, et al. Tinkering Cis Motifs Jigsaw Puzzle Led to Root-Specific Drought-Inducible Novel Synthetic Promoters. Int J Mol Sci. 2020 Feb 18;21(4).

401. Zuo J, Chua NH. Chemical-inducible systems for regulated expression of plant genes. Curr Opin Biotechnol. 2000 Apr;11(2):146–51.

402. Martínez de Alba AE, Elvira-Matelot E, Vaucheret H. Gene silencing in plants: a diversity of pathways. Biochim Biophys Acta. 2013 Dec;1829(12):1300–8.

403. Meyer P, Saedler H. Homology-dependent gene silencing in plants. Annu Rev Plant Physiol Plant Mol Biol. 1996 Jun;47:23–48.

404. Rajeevkumar S, Anunanthini P, Sathishkumar R. Epigenetic silencing in transgenic plants. Front Plant Sci. 2015 Sep 10;6:693.

405. Harpster MH, Townsend JA, Jones JD, Bedbrook J, Dunsmuir P. Relative strengths of the 35S cauliflower mosaic virus, 1', 2', and nopaline synthase promoters in transformed tobacco sugarbeet and oilseed rape callus tissue. Mol Gen Genet. 1988 Apr;212(1):182–90.

406. Huang S, An YQ, McDowell JM, McKinney EC, Meagher RB. The Arabidopsis ACT11 actin gene is strongly expressed in tissues of the emerging inflorescence, pollen, and developing ovules. Plant Mol Biol. 1997 Jan;33(1):125–39.

407. Christensen AH, Quail PH. Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. Transgenic Res. 1996 May;5(3):213–8.

408. Sanders PR, Winter JA, Barnason AR, Rogers SG, Fraley RT. Comparison of cauliflower mosaic virus 35S and nopaline synthase promoters in transgenic plants. Nucleic Acids Res. 1987 Feb 25;15(4):1543–58.

409. Sainsbury F, Thuenemann EC, Lomonossoff GP. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J. 2009 Sep;7(7):682–93.

410. Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA. A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. Plant J. 2016 Dec;88(6):1058–70.

411. Furusawa C, Kaneko K. Zipf's law in gene expression. Phys Rev Lett. 2003 Feb 28;90(8):088102.

412. Conforte AJ, Guimarães-Dias F, Neves-Borges AC, Bencke-Malato M, Felix-Whipps D, Alves-Ferreira M. Isolation and characterization of a promoter

responsive to salt, osmotic and dehydration stresses in soybean. Genet Mol Biol. 2017 Mar 27;40(1 suppl 1):226–37.

413. Sorkina A, Bardosh G, Liu Y-Z, Fridman I, Schlizerman L, Zur N, et al. Isolation of a citrus promoter specific for reproductive organs and its functional analysis in isolated juice sacs and tomato. Plant Cell Rep. 2011 Sep;30(9):1627–40.

414. Negi N, Khurana P. A salicylic acid inducible mulberry WRKY transcription factor, MiWRKY53 is involved in plant defence response. Plant Cell Rep. 2021 Nov;40(11):2151–71.

415. He Y, Zhang T, Sun H, Zhan H, Zhao Y. A reporter for noninvasively monitoring gene expression and plant transformation. Hortic Res. 2020 Sep 19;7(1):152.

416. Ricci-Tam C, Ben-Zion I, Wang J, Palme J, Li A, Savir Y, et al. Decoupling transcription factor expression and activity enables dimmer switch gene regulation. Science. 2021 Apr 16;372(6539):292–5.

417. Zrimec J, Börlin CS, Buric F, Muhammad AS, Chen R, Siewers V, et al. Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. Nat Commun. 2020 Dec 1;11(1):6141.

418. Ceroni F, Algar R, Stan G-B, Ellis T. Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods. 2015 May;12(5):415–8.

419. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature. 2015 Dec 3;528(7580):99–104.

420. Li G-W, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014 Apr 24;157(3):624–35.

421. Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013 Mar 1;339(6123):1074–7.

422. Jores T, Tonnies J, Dorrity MW, Cuperus JT, Fields S, Queitsch C. Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. Plant Cell. 2020 Jul;32(7):2120–31.

423. Jores T, Tonnies J, Wrightsman T, Buckler ES, Cuperus JT, Fields S, et al. Synthetic promoter designs enabled by a comprehensive analysis of plant core promoters. Nat Plants. 2021 Jun 3;7(6):842–55.

424. Cai Y-M, Kallam K, Tidd H, Gendarini G, Salzman A, Patron NJ. Rational design of minimal synthetic promoters for plants. Nucleic Acids Res. 2020 Dec 2;48(21):11845–56.

425. Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. Nucleic Acids Res. 2013 Dec;41(22):10668–78.

426. Ajikumar PK, Xiao W-H, Tyo KEJ, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science. 2010 Oct 1;330(6000):70–4.

427. Pfleger BF, Pitera DJ, Smolke CD, Keasling JD. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nat Biotechnol. 2006 Aug;24(8):1027–32.

428. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923–30.

429. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

430. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

431. Patron NJ, Orzaez D, Marillonnet S, Warzecha H, Matthewman C, Youles M, et al. Standards for plant synthetic biology: a common syntax for exchange of DNA parts. New Phytol. 2015 Oct;208(1):13–9.

432. Patron NJ. DNA assembly for plant biology. Curr Protoc Plant Biol. 2016 Dec;1(4):604–16.

433. Sparkes IA, Runions J, Kearns A, Hawes C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc. 2006;1(4):2019–25.

434. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol. 2000 Apr;42(6):819–32.

435. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012 Feb 14;13:31.

436. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 2011 May 2;17(1):10.

437. Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 1987 Dec 20;6(13):3901–7.

438. Thompson MG, Cruz-Morales P, Moore WM, Pearson AN, Keasling JD, Scheller HV, et al. Draft Genome Sequence of Agrobacterium fabrum ARqua1. Microbiol Resour Announc. 2020 Jul 2;9(27).

439. Hedges JB, Ryan KS. Biosynthetic Pathways to Nonproteinogenic α-Amino Acids. Chem Rev. 2020 Mar 25;120(6):3161–209.

440. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013 Jan;30(1):108–60.

441. Süssmuth RD, Mainz A. Nonribosomal Peptide Synthesis-Principles and Prospects. Angew Chem Int Ed. 2017 Mar 27;56(14):3770–821.

442. Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, et al. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science. 2012 Oct 19;338(6105):387–90.

443. Schultz AW, Lewis CA, Luzung MR, Baran PS, Moore BS. Functional characterization of the cyclomarin/cyclomarazine prenyltransferase CymD directs the biosynthesis of unnatural cyclic peptides. J Nat Prod. 2010 Mar 26;73(3):373–7.

444. Heemstra JR, Walsh CT, Sattely ES. Enzymatic tailoring of ornithine in the biosynthesis of the Rhizobium cyclic trihydroxamate siderophore vicibactin. J Am Chem Soc. 2009 Oct 28;131(42):15317–29.

445. Chekan JR, Estrada P, Covello PS, Nair SK. Characterization of the macrocyclase involved in the biosynthesis of RiPP cyclic peptides in plants. Proc Natl Acad Sci USA. 2017 Jun 20;114(25):6551–6.

446. Zwick CR, Sosa MB, Renata H. Characterization of a Citrulline 4-Hydroxylase from Nonribosomal Peptide GE81112 Biosynthesis and Engineering of Its Substrate Specificity for the Chemoenzymatic Synthesis of Enduracididine. Angew Chem Int Ed. 2019 Dec 19;58(52):18854–8.

447. Neugebauer ME, Sumida KH, Pelton JG, McMurry JL, Marchand JA, Chang MCY. A family of radical halogenases for the engineering of amino-acid-based products. Nat Chem Biol. 2019 Oct;15(10):1009–16.

448. Marchand JA, Neugebauer ME, Ing MC, Lin CI, Pelton JG, Chang MCY. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. Nature. 2019 Mar 13;567(7748):420–4.

449. Van Cura D, Ng TL, Huang J, Hager H, Hartwig JF, Keasling JD, et al. Discovery of the Azaserine Biosynthetic Pathway Uncovers a Biological Route for α-Diazoester Production. Angew Chem Int Ed. 2023 Jul 10;62(28):e202304646.

450. Neumann CS, Jiang W, Heemstra JR, Gontang EA, Kolter R, Walsh CT. Biosynthesis of piperazic acid via N5-hydroxy-ornithine in Kutzneria spp. 744. Chembiochem. 2012 May 7;13(7):972–6.

451. Wang M, Niikura H, He H-Y, Daniel-Ivad P, Ryan KS. Biosynthesis of the N-N-Bond-Containing Compound l-Alanosine. Angew Chem Int Ed. 2020 Mar 2;59(10):3881–5.

452. Shimo S, Ushimaru R, Engelbrecht A, Harada M, Miyamoto K, Kulik A, et al. Stereodivergent Nitrocyclopropane Formation during Biosynthesis of Belactosins and Hormaomycins. J Am Chem Soc. 2021 Nov 10;143(44):18413–8.

453. Bunno R, Awakawa T, Mori T, Abe I. Aziridine Formation by a FeII /α-Ketoglutarate Dependent Oxygenase and 2-Aminoisobutyrate Biosynthesis in Fungi. Angew Chem Int Ed. 2021 Jul 12;60(29):15827–31.

454. Yan F, Müller R. Class I Methyltransferase VioH Catalyzes Unusual S-Adenosyl-l-methionine Cyclization Leading to 4-Methylazetidinecarboxylic Acid Formation during Vioprolide Biosynthesis. ACS Chem Biol. 2019 Jan 18;14(1):99–105.

455. Murphy CD, Schaffrath C, O'Hagan D. Fluorinated natural products: the biosynthesis of fluoroacetate and 4-fluorothreonine in Streptomyces cattleya. Chemosphere. 2003 Jul;52(2):455–61.

456. Galván AE, Paul NP, Chen J, Yoshinaga-Sakurai K, Utturkar SM, Rosen BP, et al. Identification of the biosynthetic gene cluster for the organoarsenical antibiotic arsinothricin. Microbiol Spectr. 2021 Sep 3;9(1):e0050221.

457. Turanov AA, Xu X-M, Carlson BA, Yoo M-H, Gladyshev VN, Hatfield DL. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. Adv Nutr. 2011 Mar 10;2(2):122–8.

458. Barr I, Latham JA, Iavarone AT, Chantarojsiri T, Hwang JD, Klinman JP. Demonstration That the Radical S-Adenosylmethionine (SAM) Enzyme PqqE

Catalyzes de Novo Carbon-Carbon Cross-linking within a Peptide Substrate PqqA in the Presence of the Peptide Chaperone PqqD. J Biol Chem. 2016 Apr 22;291(17):8877–84.

459. Barondeau DP, Putnam CD, Kassmann CJ, Tainer JA, Getzoff ED. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. Proc Natl Acad Sci USA. 2003 Oct 14;100(21):12111–6.

460. Schirch V, Szebenyi DM. Serine hydroxymethyltransferase revisited. Curr Opin Chem Biol. 2005 Oct;9(5):482–7.

461. Scarsdale JN, Radaev S, Kazanina G, Schirch V, Wright HT. Crystal structure at 2.4 A resolution of E. coli serine hydroxymethyltransferase in complex with glycine substrate and 5-formyl tetrahydrofolate. J Mol Biol. 2000 Feb 11;296(1):155–68.

462. Szebenyi DME, Musayev FN, di Salvo ML, Safo MK, Schirch V. Serine hydroxymethyltransferase: role of glu75 and evidence that serine is cleaved by a retroaldol mechanism. Biochemistry. 2004 Jun 8;43(22):6865–76.

463. Xu Y, Tao F, Ma C, Xu P. New constitutive vectors: useful genetic engineering tools for biocatalysis. Appl Environ Microbiol. 2013 Apr;79(8):2836–40.

464. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods. 2015 Jun;12(6):523–6.

465. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

466. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022 Jun;19(6):679–82.

467. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31(2):455–61.

468. Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. 2016 May;11(5):905–19.

469. Tricot C, Vander Wauven C, Wattiez R, Falmagne P, Stalon V. Purification and properties of a succinyltransferase from Pseudomonas aeruginosa specific for both arginine and ornithine. Eur J Biochem. 1994 Sep 15;224(3):853–61.

470. Mehmood N, Saeed M, Zafarullah S, Hyder S, Rizvi ZF, Gondal AS, et al. Multifaceted Impacts of Plant-Beneficial Pseudomonas spp. in Managing Various Plant Diseases and Crop Yield Improvement. ACS Omega. 2023 Jun 27;8(25):22296–315.

471. Adams DO, Yang SF. Ethylene biosynthesis: Identification of 1-aminocyclopropane-1-carboxylic acid as an intermediate in the conversion of methionine to ethylene. Proc Natl Acad Sci USA. 1979 Jan;76(1):170–4.

472. Mersmann S, Bourdais G, Rietz S, Robatzek S. Ethylene signaling regulates accumulation of the FLS2 receptor and is required for the oxidative burst contributing to plant immunity. Plant Physiol. 2010 Sep;154(1):391–400.

473. van Loon LC, Geraats BPJ, Linthorst HJM. Ethylene as a modulator of disease resistance in plants. Trends Plant Sci. 2006 Apr;11(4):184–91.

474. Gamalero E, Glick BR. Bacterial modulation of plant ethylene levels. Plant Physiol. 2015 Sep;169(1):13–22.

475. Ma S, Mandalapu D, Wang S, Zhang Q. Biosynthesis of cyclopropane in natural products. Nat Prod Rep. 2022 May 26;39(5):926–45.

476. Talele TT. The “Cyclopropyl Fragment” is a Versatile Player that Frequently Appears in Preclinical/Clinical Drug Molecules. J Med Chem. 2016 Oct 13;59(19):8712–56.

477. Souillart L, Cramer N. Catalytic C-C Bond Activations via Oxidative Addition to Transition Metals. Chem Rev. 2015 Sep 9;115(17):9410–64.

478. Iwig DF, Uchida A, Stromberg JA, Booker SJ. The activity of Escherichia coli cyclopropane fatty acid synthase depends on the presence of bicarbonate. J Am Chem Soc. 2005 Aug 24;127(33):11612–3.

479. Roach CR, Hall DE, Zerbe P, Bohlmann J. Plasticity and evolution of (+)-3-carene synthase and (-)-sabinene synthase functions of a sitka spruce monoterpene synthase gene family associated with weevil resistance. J Biol Chem. 2014 Aug 22;289(34):23859–69.

480. Jakubczyk D, Caputi L, Hatsch A, Nielsen CAF, Diefenbacher M, Klein J, et al. Discovery and reconstitution of the cycloclavine biosynthetic pathway--enzymatic formation of a cyclopropyl group. Angew Chem Int Ed. 2015 Apr 20;54(17):5117–21.

481. Neumann CS, Walsh CT. Biosynthesis of (-)-(1S,2R)-allocoronamic acyl thioester by an Fe(II)-dependent halogenase and a cyclopropane-forming flavoprotein. J Am Chem Soc. 2008 Oct 29;130(43):14022–3.

482. Bell EA, Qureshi MY, Pryce RJ, Janzen DH, Lemke P, Clardy J. 2,4-Methanoproline (2-carboxy-2,4-methanopyrrolidine) and 2,4-methanoglutamic acid (1-amino-1,3-dicarboxycyclobutane) in seeds of Ateleia herbert smithii Pittier (Leguminosae). J Am Chem Soc. 1980 Feb;102(4):1409–12.

483. Hong YJ, Tantillo DJ. How cyclobutanes are assembled in nature--insights from quantum chemistry. Chem Soc Rev. 2014 Jul 21;43(14):5042–50.

484. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009 May;6(5):343–5.

485. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403–10.

486. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim Biophys Acta. 2015 Aug;1854(8):1019–37.

487. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.

488. Hmelo LR, Borlee BR, Almblad H, Love ME, Randall TE, Tseng BS, et al. Precision-engineering the *Pseudomonas aeruginosa* genome with two-step allelic exchange. Nat Protoc. 2015 Nov;10(11):1820–41.

489. Gaussian. 2016.

490. Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A Gen Phys. 1988 Sep 15;38(6):3098–100.

491. Li J, Sha Y. A convenient synthesis of amino acid methyl esters. Molecules. 2008 May 8;13(5):1111–9.

492. Anderson CT. We be jammin': an update on pectin biosynthesis, trafficking and dynamics. J Exp Bot. 2016 Jan;67(2):495–502.

493. Wachananawat B, Kuroha T, Takenaka Y, Kajiura H, Naramoto S, Yokoyama R, et al. Diversity of pectin rhamnogalacturonan I rhamnosyltransferases in glycosyltransferase family 106. Front Plant Sci. 2020 Jul 2;11:997.

494. Atmodjo MA, Hao Z, Mohnen D. Evolving views of pectin biosynthesis. Annu Rev Plant Biol. 2013 Mar 1;64:747–79.

495. Loqué D, Scheller HV, Pauly M. Engineering of plant cell walls for enhanced biofuel production. Curr Opin Plant Biol. 2015 Jun 3;25:151–61.

496. Takenaka Y, Kato K, Ogawa-Ohnishi M, Tsuruhama K, Kajiura H, Yagyu K, et al. Pectin RG-I rhamnosyltransferases represent a novel plant-specific glycosyltransferase family. Nat Plants. 2018 Sep;4(9):669–76.

497. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature. 2017 Apr 6;544(7648):65–70.

498. Lau JM, McNeil M, Darvill AG, Albersheim P. Structure of the backbone of rhamnogalacturonan I, a pectic polysaccharide in the primary cell walls of plants. Carbohydr Res. 1985 Mar;137:111–25.

499. Lau JM, McNeil M, Darvill AG, Albersheim P. Treatment of rhamnogalacturonan I with lithium in ethylenediamine. Carbohydr Res. 1987 Nov;168(2):245–74.

500. Uehara Y, Tamura S, Maki Y, Yagyu K, Mizoguchi T, Tamiaki H, et al. Biochemical characterization of rhamnosyltransferase involved in biosynthesis of pectic rhamnogalacturonan I in plant cell wall. Biochem Biophys Res Commun. 2017 Apr 22;486(1):130–6.

501. Kozlova LV, Nazipova AR, Gorshkov OV, Petrova AA, Gorshkova TA. Elongating maize root: zone-specific combinations of polysaccharides from type I and type II primary cell walls. Sci Rep. 2020 Jul 2;10(1):10956.

502. Kamal N, Mun T, Reid D, Lin J-S, Akyol TY, Sandal N, et al. Insights into the evolution of symbiosis gene copy number and distribution from a chromosome-scale Lotus japonicus Gifu genome sequence. DNA Res. 2020 Jun 1;27(3).

503. Mun T, Bachmann A, Gupta V, Stougaard J, Andersen SU. Lotus Base: An integrated information portal for the model legume Lotus japonicus. Sci Rep. 2016 Dec 23;6:39447.

504. Małolepszy A, Mun T, Sandal N, Gupta V, Dubin M, Urbański D, et al. The LORE1 insertion mutant resource. Plant J. 2016 Sep 27;88(2):306–17.

505. Egelund J, Obel N, Ulvskov P, Geshi N, Pauly M, Bacic A, et al. Molecular characterization of two Arabidopsis thaliana glycosyltransferase mutants, rra1 and rra2, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. Plant Mol Biol. 2007 Jul;64(4):439–51.

506. Scavuzzo-Duggan T, Varoquaux N, Madera M, Vogel JP, Dahlberg J, Hutmacher R, et al. Cell Wall Compositions of Sorghum bicolor Leaves and Roots Remain Relatively Constant Under Drought Conditions. Front Plant Sci. 2021 Nov 12;12:747225.

507. Stonebloom S, Ebert B, Xiong G, Pattathil S, Birdseye D, Lao J, et al. A DUF-246 family glycosyltransferase-like gene affects male fertility and the biosynthesis of pectic arabinogalactans. BMC Plant Biol. 2016 Apr 18;16:90.

508. Harholt J, Jensen JK, Sørensen SO, Orfila C, Pauly M, Scheller HV. ARABINAN DEFICIENT 1 is a putative arabinosyltransferase involved in biosynthesis of pectic arabinan in Arabidopsis. Plant Physiol. 2006 Jan;140(1):49–58.

509. Ebert B, Birdseye D, Liwanag AJM, Laursen T, Rennie EA, Guo X, et al. The Three Members of the Arabidopsis Glycosyltransferase Family 92 are Functional β-1,4-Galactan Synthases. Plant Cell Physiol. 2018 Dec 1;59(12):2624–36.

510. Saffer AM, Carpita NC, Irish VF. Rhamnose-Containing Cell Wall Polymers Suppress Helical Plant Growth Independently of Microtubule Orientation. Curr Biol. 2017 Aug 7;27(15):2248-2259.e4.

511. Iijima M, Higuchi T, Barlow PW. Contribution of root cap mucilage and presence of an intact root cap in maize (Zea mays) to the reduction of soil mechanical impedance. Ann Bot. 2004 Sep;94(3):473–7.

512. Gao Y, Lipton AS, Munson CR, Ma Y, Johnson KL, Murray DT, et al. Elongated galactan side chains mediate cellulose-pectin interactions in engineered Arabidopsis secondary cell walls. Plant J. 2023 Jul;115(2):529–45.

513. Jones P, Messner B, Nakajima J-I, Schäffner AR, Saito K. UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J Biol Chem. 2003 Nov 7;278(45):43910–8.

514. Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, et al. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell. 2020 Jan;32(1):15–41.

515. Stateczny D, Oppenheimer J, Bommert P. G protein signaling in plants: minus times minus equals plus. Curr Opin Plant Biol. 2016 Dec;34:127–35.

516. Maruta N, Trusov Y, Jones AM, Botella JR. Heterotrimeric G proteins in plants: canonical and atypical gα subunits. Int J Mol Sci. 2021 Oct 31;22(21).

517. Chakravorty D, Gookin TE, Milner MJ, Yu Y, Assmann SM. Extra-Large G Proteins Expand the Repertoire of Subunits in Arabidopsis Heterotrimeric G Protein Signaling. Plant Physiol. 2015 Sep;169(1):512–29.

518. Cui Y, Jiang N, Xu Z, Xu Q. Heterotrimeric G protein are involved in the regulation of multiple agronomic traits and stress tolerance in rice. BMC Plant Biol. 2020 Feb 28;20(1):90.

519. Bovin AD, Pavlova OA, Dolgikh AV, Leppyanen IV, Dolgikh EA. The Role of Heterotrimeric G-Protein Beta Subunits During Nodulation in Medicago truncatula Gaertn and Pisum sativum L. Front Plant Sci. 2021;12:808573.

520. Choudhury SR, Pandey S. Specific subunits of heterotrimeric G proteins play important roles during nodulation in soybean. Plant Physiol. 2013 May;162(1):522–33.

521. Choudhury SR, Pandey S. Phosphorylation-Dependent Regulation of G-Protein Cycle during Nodule Formation in Soybean. Plant Cell. 2015 Nov;27(11):3260–76.

522. Urano D, Chen J-G, Botella JR, Jones AM. Heterotrimeric G protein signalling in the plant kingdom. Open Biol. 2013 Mar 27;3(3):120186.

523. Yi T-M, Kitano H, Simon MI. A quantitative characterization of the yeast heterotrimeric G protein cycle. Proc Natl Acad Sci USA. 2003 Sep 16;100(19):10764–9.

524. Maruta N, Trusov Y, Brenya E, Parekh U, Botella JR. Membrane-localized extra-large G proteins and Gbg of the heterotrimeric G proteins form functional complexes engaged in plant immunity in Arabidopsis. Plant Physiol. 2015 Mar;167(3):1004–16.

525. Ding L, Pandey S, Assmann SM. Arabidopsis extra-large G proteins (XLGs) regulate root morphogenesis. Plant J. 2008 Jan;53(2):248–63.

526. MacLean AM, Bravo A, Harrison MJ. Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. Plant Cell. 2017 Oct;29(10):2319–35.

527. Lepetit M, Brouquisse R. Control of the rhizobium-legume symbiosis by the plant nitrogen demand is tightly integrated at the whole plant level and requires inter-organ systemic signaling. Front Plant Sci. 2023 Mar 9;14:1114840.

528. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018 Jun 1;35(6):1547–9.

529. Peter Henry Andrews Sneath, Sokal RR. Numerical Taxonomy: The Principles and Practice of Numerical Classification. 1973;

530. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985 Jul;39(4):783–91.

531. Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. J Theor Biol. 1965 Mar;8(2):357–66.

532. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389–402.

533. Pearson AN, Thompson MG, Kirkpatrick LD, Ho C, Vuu KM, Waldburger LM, et al. The pGinger Family of Expression Plasmids. Microbiol Spectr. 2023 Jun 15;11(3):e0037323.

534. Zhou A, Kirkpatrick LD, Ornelas IJ, Washington LJ, Hummel NFC, Gee CW, et al. A suite of constitutive promoters for tuning gene expression in plants. ACS Synth Biol. 2023 May 19;12(5):1533–45.

535. Schindler U, Beckmann H, Cashmore AR. HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. Plant J. 1993 Jul;4(1):137–50.

536. Genre A, Russo G. Does a Common Pathway Transduce Symbiotic Signals in Plant-Microbe Interactions? Front Plant Sci. 2016 Feb 16;7:96.

537. Carrere S, Verdier J, Gamas P. MtExpress, a Comprehensive and Curated RNAseq-based Gene Expression Atlas for the Model Legume Medicago truncatula. Plant Cell Physiol. 2021 Nov 17;62(9):1494–500.

538. Liang Y, Gao Y, Jones AM. Extra Large G-Protein Interactome Reveals Multiple Stress Response Function and Partner-Dependent XLG Subcellular Localization. Front Plant Sci. 2017 Jun 13;8:1015.

539. Drew MC. Comparison of the effects of a localised supply of phosphate, nitrate, ammonium and potassium on the growth of the seminal root system, and the shoot, in barley. New Phytol. 1975 Nov;75(3):479–90.

540. Krapp A, Berthomé R, Orsel M, Mercey-Boutet S, Yu A, Castaings L, et al. Arabidopsis roots and shoots show distinct temporal adaptation patterns toward nitrogen starvation. Plant Physiol. 2011 Nov;157(3):1255–82.

541. Vernoud V, Horton AC, Yang Z, Nielsen E. Analysis of the small GTPase gene superfamily of Arabidopsis. Plant Physiol. 2003 Mar;131(3):1191–208.

542. Mohanasundaram B, Dodds A, Kukshal V, Jez JM, Pandey S. Distribution and the evolutionary history of G-protein components in plant and algal lineages. Plant Physiol. 2022 Jun 27;189(3):1519–35.

543. Cantos CF, dePamphilis CW, Assmann SM. Extra-large G proteins have extra-large effects on agronomic traits and stress tolerance in maize and rice. Trends Plant Sci. 2023 Sep;28(9):1033–44.

544. Urano D, Maruta N, Trusov Y, Stoian R, Wu Q, Liang Y, et al. Saltational evolution of the heterotrimeric G protein signaling mechanisms in the plant kingdom. Sci Signal. 2016 Sep 20;9(446):ra93.

545. Kalinina NO, Makarova S, Makhotenko A, Love AJ, Taliansky M. The multiple functions of the nucleolus in plant development, disease and stress responses. Front Plant Sci. 2018 Feb 9;9:132.

546. Quan W, Chan Z, Wei P, Mao Y, Bartels D, Liu X. PHD finger proteins function in plant development and abiotic stress responses: an overview. Front Plant Sci. 2023 Nov 17;14:1297607.

547. Heo JB, Sung S, Assmann SM. Ca2+-dependent GTPase, extra-large G protein 2 (XLG2), promotes activation of DNA-binding protein related to vernalization 1 (RTV1), leading to activation of floral integrator genes and early flowering in Arabidopsis. J Biol Chem. 2012 Mar 9;287(11):8242–53.

548. Liang X, Ding P, Lian K, Wang J, Ma M, Li L, et al. Arabidopsis heterotrimeric G proteins regulate immunity by directly coupling to the FLS2 receptor. eLife. 2016 Apr 4;5:e13568.

549. Bhardwaj D, Samantaray JP, Kour V, Ganotra J, Verma R, Chaubey A, et al. Pea G-protein γ subunits: Unlocking their potential in physiological stress and mycorrhizal mediated nutrient sensing. Current Plant Biology. 2024 Jun;38:100344.

550. Liu J, Liu MX, Qiu LP, Xie F. SPIKE1 Activates the GTPase ROP6 to Guide the Polarized Growth of Infection Threads in Lotus japonicus. Plant Cell. 2020 Dec;32(12):3774–91.

551. Lei M-J, Wang Q, Li X, Chen A, Luo L, Xie Y, et al. The small GTPase ROP10 of Medicago truncatula is required for both tip growth of root hairs and nod factor-induced root hair deformation. Plant Cell. 2015 Mar 20;27(3):806–22.

552. Engelhardt S, Trutzenberg A, Hückelhoven R. Regulation and Functions of ROP GTPases in Plant-Microbe Interactions. Cells. 2020 Sep 2;9(9).

553. Frank M, Fechete LI, Tedeschi F, Nadzieja M, Nørgaard MMM, Montiel J, et al. Single-cell analysis identifies genes facilitating rhizobium infection in Lotus japonicus. Nat Commun. 2023 Nov 7;14(1):7171.

554. Serrano K, Bezrutczyk M, Goudeau D, Dao T, O'Malley R, Malmstrom RR, et al. Spatial co-transcriptomics reveals discrete stages of the arbuscular mycorrhizal symbiosis. Nat Plants. 2024 Apr 8;10(4):673–88.

555. Roux B, Rodde N, Jardinaud M-F, Timmers T, Sauviac L, Cottret L, et al. An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. Plant J. 2014 Mar;77(6):817–37.

556. Schiessl K, Lilley JLS, Lee T, Tamvakis I, Kohlen W, Bailey PC, et al. NODULE INCEPTION Recruits the Lateral Root Developmental Program for Symbiotic Nodule Organogenesis in Medicago truncatula. Curr Biol. 2019 Nov 4;29(21):3657-3668.e5.

557. Karlo M, Boschiero C, Landerslev KG, Blanco GS, Wen J, Mysore KS, et al. The CLE53-SUNN genetic pathway negatively regulates arbuscular mycorrhiza root colonization in Medicago truncatula. J Exp Bot. 2020 Aug 6;71(16):4972–84.

558. Jardinaud M-F, Boivin S, Rodde N, Catrice O, Kisiala A, Lepage A, et al. A Laser Dissection-RNAseq Analysis Highlights the Activation of Cytokinin Pathways by Nod Factors in the Medicago truncatula Root Epidermis. Plant Physiol. 2016 Jul;171(3):2256–76.

559. Zeng T, Holmer R, Hontelez J, Te Lintel-Hekkert B, Marufu L, de Zeeuw T, et al. Host- and stage-dependent secretome of the arbuscular mycorrhizal fungus Rhizophagus irregularis. Plant J. 2018 May;94(3):411–25.

560. Pereira WJ, Boyd J, Conde D, Triozzi PM, Balmant KM, Dervinis C, et al. The single-cell transcriptome program of nodule development cellular lineages in Medicago truncatula. Cell Rep. 2024 Feb 27;43(2):113747.

561. Marie C. Abnormal Bacteroid Development in Nodules Induced by a Glucosamine Synthase Mutant of *Rhizobium ieguminosarum*. MPMI. 1994;7(4):482.

562. Malolepszy A, Kelly S, Sørensen KK, James EK, Kalisch C, Bozsoki Z, et al. A plant chitinase controls cortical infection thread progression and nitrogen-fixing symbiosis. eLife. 2018 Oct 4;7.

563. Pumplin N, Harrison MJ. Live-cell imaging reveals periarbuscular membrane domains and organelle location in Medicago truncatula roots during arbuscular mycorrhizal symbiosis. Plant Physiol. 2009 Oct;151(2):809–19.

Appendices

Supplemental Table 3-1. Primers used in study

Primer ID	Primer Sequence	Use
LjUBQ Ref Fwd	GGTCGAAAGCTCTGACACTATT	qPCR
LjUBQ Ref Rev	CCTCAAGGGTGATGGTCTTG	qPCR
LjARAD1 qPCR Fwd	CTGCTGCTATAAAGCCAGGATAT	qPCR
LjARAD1 qPCR Rev	GCGCCAAATCTCATTAACCG	qPCR
LjGALS1 qPCR Fwd	CTCTGCCTCTCTTCTGATTATGG	qPCR
LjGALS1 qPCR Rev	TGCTATGGTTCCATGGTAGTG	qPCR
P2	CCATGGCGGTTCCGTGAATCTTAGG	Genotyping
rrt1-1 Fwd	CCCCCACATGTCACCCAAAGCAGT	Genotyping
rrt1-1 Rev	GTGGCAGGAGATCTCGCCGTATCG	Genotyping
rrt1-2 Fwd	GCGGCATGCTGTAAAGTAACCCGCTT	Genotyping
rrt1-2 Rev	TGGACTTTGGATTGTGTTGCTAGTGGGG	Genotyping

Supplemental Table 3-2. Raw log-fold ddCt values for qPCR

Treatmen	Sample	Gene	log(2^-ΔΔCt) relative to LjUB
rrt1-1	1_1	*LjARAD1*	0.961341865
rrt1-1	1_2	*LjARAD1*	1.117931255
rrt1-1	1_3	*LjARAD1*	1.204262098
rrt1-1	2_1	*LjARAD1*	0.86484306
rrt1-1	2_2	*LjARAD1*	0.833021446
rrt1-1	2_3	*LjARAD1*	0.8989495
rrt1-1	3_1	*LjARAD1*	0.141955241
rrt1-1	3_2	*LjARAD1*	0.802279686
rrt1-1	3_3	*LjARAD1*	0.809188924
Gifu	1_1	*LjARAD1*	1.353129092
Gifu	1_2	*LjARAD1*	1.076171959
Gifu	1_3	*LjARAD1*	0.775187859
Gifu	2_1	*LjARAD1*	0.220239424
Gifu	2_2	*LjARAD1*	1.046663163
Gifu	2_3	*LjARAD1*	1.189466642
Gifu	3_1	*LjARAD1*	0.415558731
Gifu	3_2	*LjARAD1*	1.340279671
Gifu	3_3	*LjARAD1*	1.583303458
rrt1-1	1_1	*LjGALS1*	1.012684922
rrt1-1	1_2	*LjGALS1*	1.078013857
rrt1-1	1_3	*LjGALS1*	1.10743738
rrt1-1	2_1	*LjGALS1*	0.736442854
rrt1-1	2_2	*LjGALS1*	0.860487915
rrt1-1	2_3	*LjGALS1*	0.711666924
rrt1-1	3_1	*LjGALS1*	0.66659286
rrt1-1	3_2	*LjGALS1*	0.664640631
rrt1-1	3_3	*LjGALS1*	0.845468081
Gifu	1_1	*LjGALS1*	1.201389
Gifu	1_2	*LjGALS1*	1.216670614
Gifu	1_3	*LjGALS1*	0.94959275
Gifu	2_1	*LjGALS1*	0.953789338
Gifu	2_2	*LjGALS1*	1.038759384
Gifu	2_3	*LjGALS1*	0.795710879
Gifu	3_1	*LjGALS1*	0.898927386
Gifu	3_2	*LjGALS1*	1.043224221
Gifu	3_3	*LjGALS1*	0.901936428

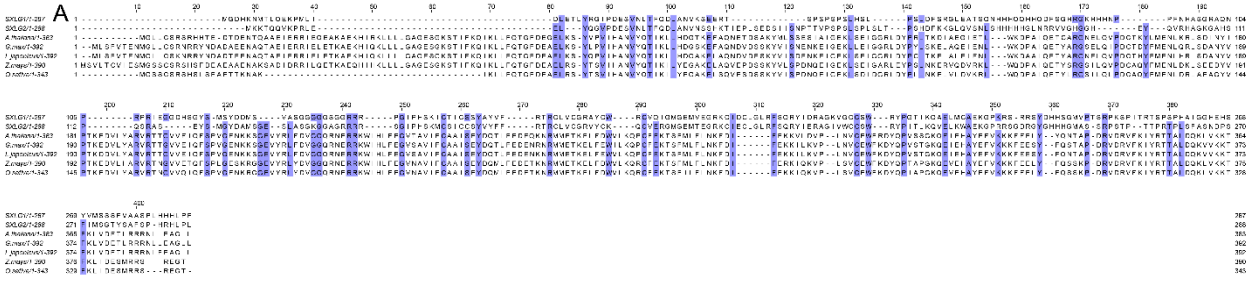
Supplemental Table 4-1. Primers used in study

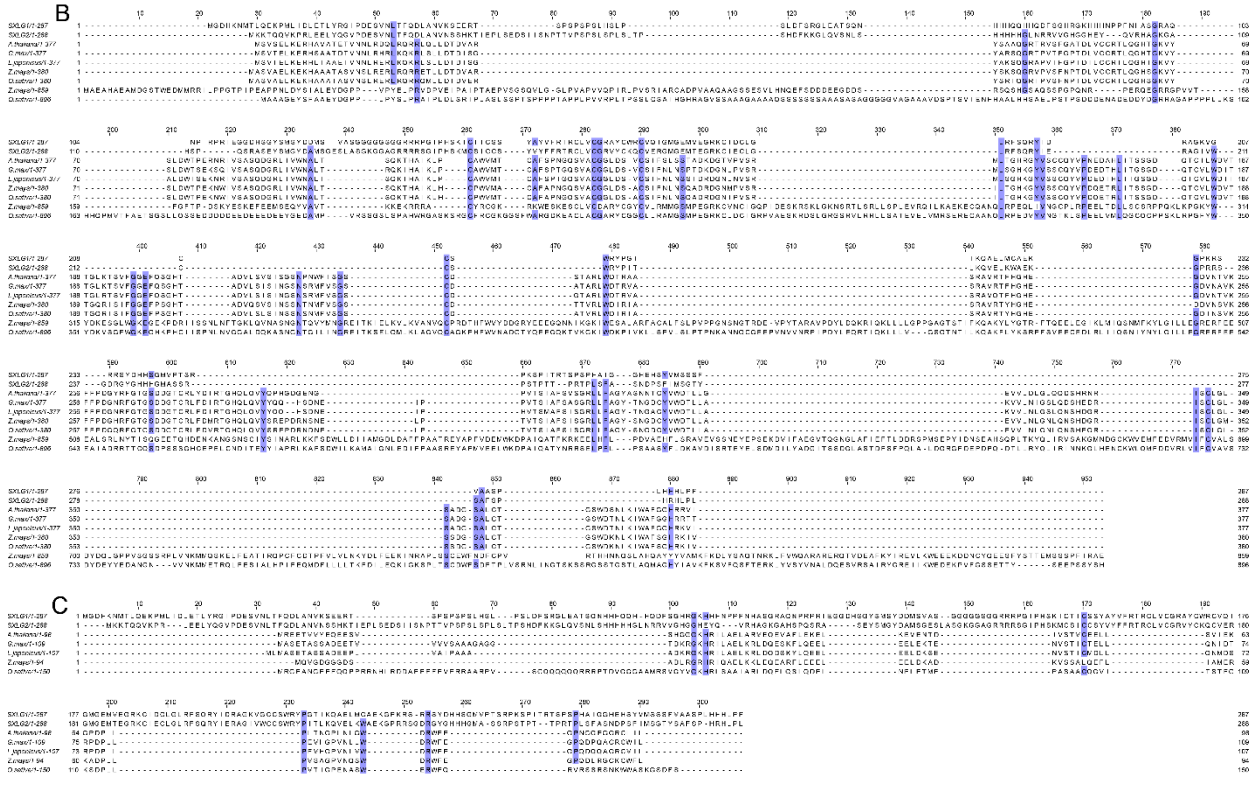
Primer ID	Primer Sequence 5'-3'	Use
LjUBQ Ref Fwd	GGTCGAAAGCTCTGACACTATT	qPCR
LjUBQ Ref Rev	CCTCAAGGGTGATGGTCTTG	qPCR
LjSXLG1 qPCR Fwd	GGCCAAAGAGAAGCAGAAGA	qPCR
LjSXLG1 qPCR Rev	CTTGTTCGGGTGATAGGACTTT	qPCR
LjSXLG2 qPCR Fwd	GGAGGACACGAATATCAGGTTAG	qPCR
LjSXLG2 qPCR Rev	TCACCACTCATGGCATCATAG	qPCR
P2	CCATGGCGGTTCCGTGAATCTTAGG	Genotyping
sxlg1 Fwd	ACCACCTCCTCCCGAAGCCACACT	Genotyping
sxlg1 Rev	TCGACCGGCCTTAGAATTGCCACA	Genotyping
sxlg2 Fwd	CCCATGCCTCTTTCCACACACTGC	Genotyping
sxlg2 Rev	CACCTTTCAAGACCTTGCCAATGTGAA	Genotyping
mUAV Backbone Fwd & Rv	tgccacctgacgtctaagaa	Split destination plasmid to increase amplification for Gibson
mUAV-SXLG2 C-term Fwd	TCGGTCTCTAATGAAGAAGACTCAGCAGG	Gibson to make SXLG2::mNG flourescent fusion protein
mUAV-SXLG2 C-term Rev	CTGAGTCTTCTTCATAGAGaccgaattcc	""
SXLG2::mNG C-term Rev	CTGACagatccaccacctccAAGAGGGAGGTGACG	""
SXLG2::mNG C-term Fwd	ggaggtggtggatctGTCAGTAAAGGAGAAG	""
SXLG::mNG C-term GG Amp	agggtCTCAAAGCTCAC	Amplify insert to improve efficiency - pairs mUAV-SXLG2 C-term Fwd
mNG::SXLG2 N-term Fwd	CAAGggaggtggtggatctATGAAGAAGACTCAGC	Gibson to make mNG::SXLG2 flourescent fusion protein
mNG::SXLG2 N-term Rev	agatccaccacctccCTTGTATAACTCATCCA	""
SXLG2-mUAV N-term Fwd	CACCTCCCTCTTTGAGCTTTGAGaccct	""
SXLG-mUAV N-term Rev	gcagggtCTCAAAGCTCAAAGAGGGAGGT	""
mUAV-SXLG1 C-term Fwd	atttctggaattcggtCTCTATGGGTGATCACAAGA	Gibson to make SXLG1::mNG flourescent fusion protein
mUAV-SXLG1 C-term Rev	CTTGTGATCACCCATAGAGaccgaattcc	""
SXLG1::mNG C-term Fwd	ggaggtggtggatctGTCAGTAAAGGAGAAG	""
SXLG1::mNG C-term Rev	ctgacagatccaccacctccAAATGGGAGGTGGTG	""
mNG::SXLG1 N-term Fwd	GGGAATGGATGAGTTATACAAGggaggtggtggatctA	Gibson to make mNG::SXLG1 flourescent fusion protein
mNG::SXLG1 N-term Rev	GTTCTTGTGATCACCCATagatccaccacctccCTT	""
SXLG1-mUAV N-term Fwd	CTCTCCTCTTCACCACCACCTCCCATTTTAAGCTTTGAG	""
SXLG1-mUAV N-term Rev	gccggactgcagggtCTCAAAGCTTAAAATGGGAGG	""

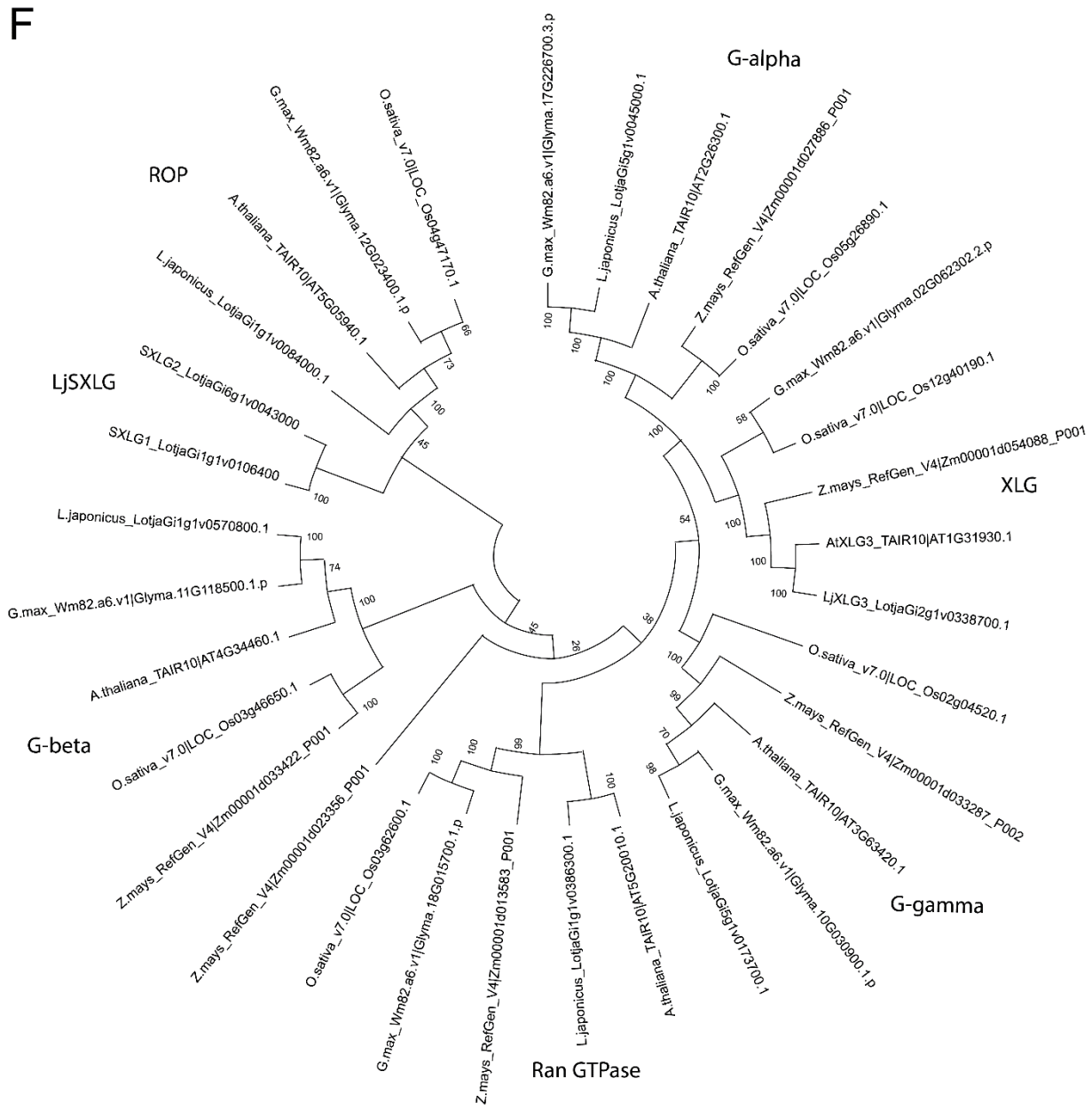
Supplemental Table 4-2. Raw log-fold ddCt values for qPCR

Treatmen	Sample	Gene	log(2^-ΔΔCt) relative to LjUBQ
sxlg2	1_1	*LjSXLG1*	-0.88924662
sxlg2	1_2	*LjSXLG1*	-1.439863943
sxlg2	1_3	*LjSXLG1*	0.1646334
sxlg2	2_1	*LjSXLG1*	0.311741992
sxlg2	2_2	*LjSXLG1*	0.298735974
sxlg2	2_3	*LjSXLG1*	0.362161927
sxlg2	3_1	*LjSXLG1*	1.242418752
sxlg2	3_2	*LjSXLG1*	1.164158951
sxlg2	3_3	*LjSXLG1*	1.57993786
Gifu	1_1	*LjSXLG1*	0.721451466
Gifu	1_2	*LjSXLG1*	0.500059496
Gifu	1_3	*LjSXLG1*	0.709015077
Gifu	2_1	*LjSXLG1*	1.165545915
Gifu	2_2	*LjSXLG1*	1.100179086
Gifu	2_3	*LjSXLG1*	1.446924002
Gifu	3_1	*LjSXLG1*	1.772923325
Gifu	3_2	*LjSXLG1*	0.774429016
Gifu	3_3	*LjSXLG1*	0.809472616
sxlg2	1_1	*LjSXLG2*	-1.579584099
sxlg2	1_2	*LjSXLG2*	-3.299135425
sxlg2	1_3	*LjSXLG2*	-1.634443531
sxlg2	2_1	*LjSXLG2*	-1.349229901
sxlg2	2_2	*LjSXLG2*	-3.555046122
sxlg2	2_3	*LjSXLG2*	-1.496880123
sxlg2	3_1	*LjSXLG2*	0.760210442
sxlg2	3_2	*LjSXLG2*	0.867877382
sxlg2	3_3	*LjSXLG2*	1.202542896
Gifu	1_1	*LjSXLG2*	0.865146621
Gifu	1_2	*LjSXLG2*	0.430516966
Gifu	1_3	*LjSXLG2*	0.657062642
Gifu	2_1	*LjSXLG2*	1.212531704
Gifu	2_2	*LjSXLG2*	1.103650931
Gifu	2_3	*LjSXLG2*	1.536709442
Gifu	3_1	*LjSXLG2*	1.372680699
Gifu	3_2	*LjSXLG2*	0.845735933
Gifu	3_3	*LjSXLG2*	0.975965062

Supplemental Figure 4-1. Alignments of SXLGs to other G-protein subunits and collective phylogenetic tree

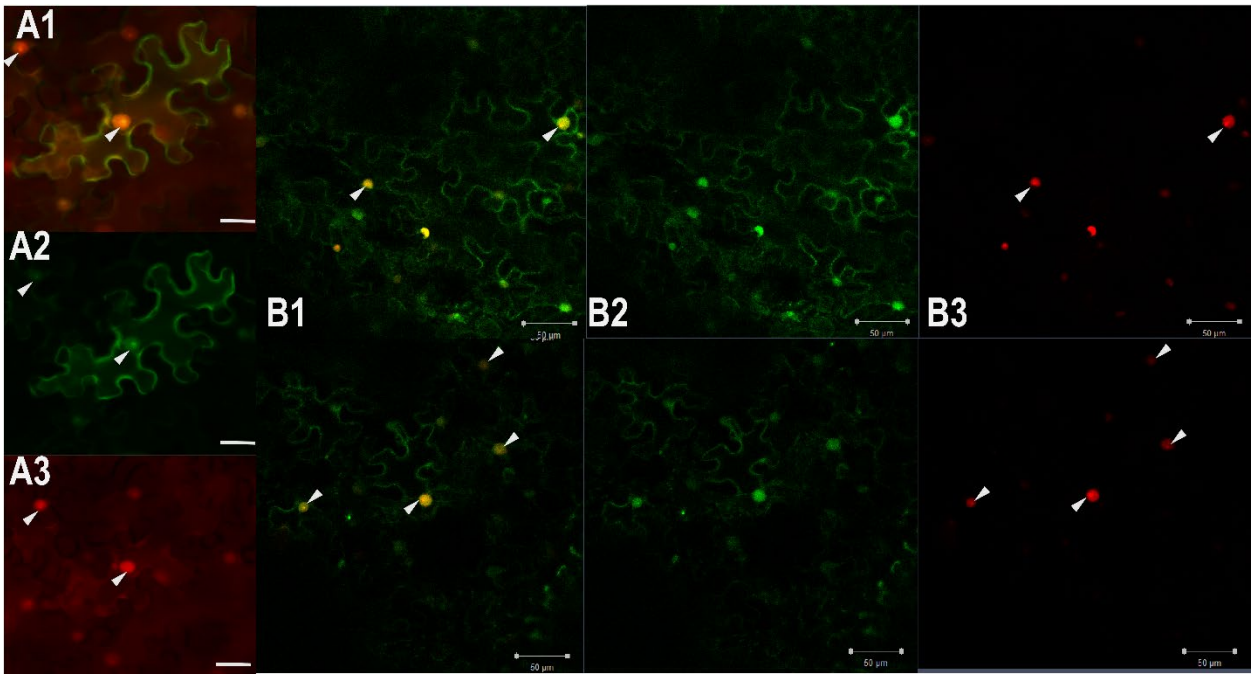






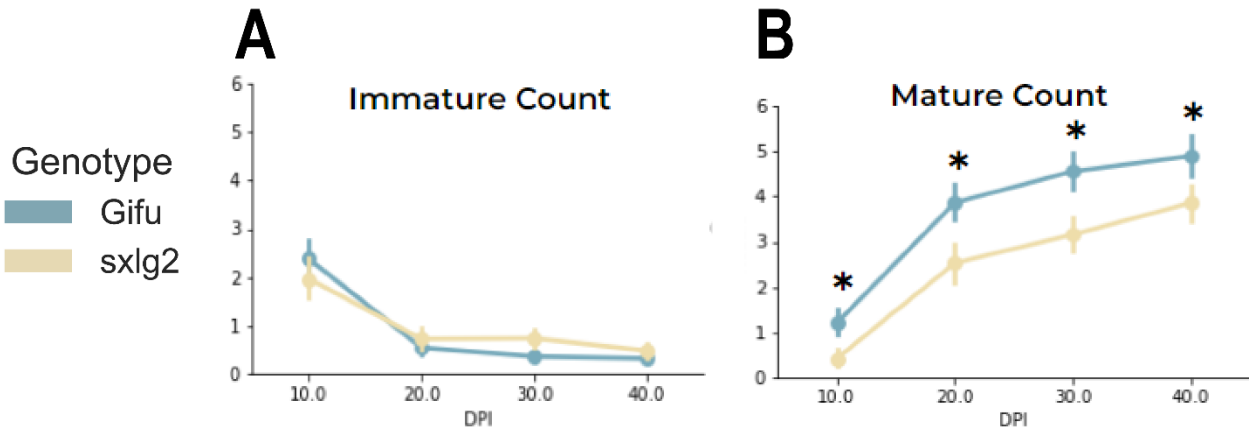
Supplemental Figure 4-1. Peptide sequence alignments of LjSXLGs with representatives of **A)** Gα subunits **B)** Gβ subunits **C)** G**γ** subunits **D)** Ran GTPases **E)** ROP GTPases from *Arabidopsis thaliana, Glycine max, Lotus japonicus, Zea mays, and Oryza sativa*. **F)** Phylogenetic tree with all of the proteins showing the distinct groupings. The gene IDs are from Phytozome.

Supplemental Figure 4-2. Transient Expression of N-terminal FP fusion of SXLG2 and free mNeonGreen



Supplemental Figure 4-2. Cells are transfected with a plasmid containing **A)** mNeonGreen (mNG) fused to the N-terminus of the coding sequence for *LjSXLG2*, driven by the pCM2 promoter. An additional plasmid containing mScarlet flanked by a nuclear localization signal is utilized as a nuclear marker. **A1)** Overlay **A2)** mNeonGreen::SXLG2 fusion **A3)** Nuclear marker. **B)** mNeonGreen expressed without a fusion, driven by pCM2 promoter. **B1)** Overlay **B2)** mNeonGreen **B3)** Nuclear marker. Results were consistent across promoter strength and infiltration OD. Nucleoli are marked by white arrows. Bars = 50uM. All images are 400x magnification.

Supplemental Figure 4-3. Non-destructive observations separated by (im)mature nodule status



Supplemental Figure 4-3. Comparison of **A)** immature & **B)** mature nodule counts between WT Gifu & Lj*sxlg2* plants grown in nitrogen depleted conditions on plates that allowed for non-destructive sample collection. P-value ≦ 0.02 denoted by *. Calculated by Kruskal-Willis Test. n = 70