Smiling in the Face of Adversity: Molecular and Evolutionary Mechanisms Behind Copper Tolerance in *Mimulus guttatus*

by

Arianti Rojas Carvajal

Department of Biology
Duke University

Defense Date: July 31st, 2024

Approved:

John Willis, Advisor

Gregory Wray

Sheng Yang-He

Robert Franks

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Biology in The Graduate School of
Duke University
2024

ABSTRACT

Smiling in the Face of Adversity: Molecular and Evolutionary Mechanisms Behind Copper Tolerance in *Mimulus guttatus*

by

Arianti Rojas Carvajal

Department of Biology
Duke University

Defense Date: July 31st, 2024

Approved:

John Willis, Advisor

Gregory Wray

Sheng Yang-He

Robert Franks

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Biology in The Graduate School of Duke University
2024

Abstract

This study investigates the molecular and evolutionary mechanisms underlying copper tolerance in *Mimulus guttatus*. I focused on the role of multi-copper oxidases (MCOs), specifically the T-MCO gene located in the Tol1 locus in *Mimulus guttatus*, to provide direct evidence of its contribution to the plant's copper tolerance mechanism. The T-MCO gene was overexpressed in non-tolerant backgrounds of *M. guttatus* and *Arabidopsis thaliana*, with results indicating that T-MCO is central to copper tolerance in *M. guttatus*. While MCO-centered mechanisms were previously associated primarily with bacteria and fungi, this finding suggests a novel copper tolerance mechanism in plants. Transgenic *A. thaliana* lines exhibited enhanced germination rates and root growth across various copper concentrations, confirming the significant impact of T-MCO overexpression. I used generalized linear mixed models (GLMMs) to validate these findings, highlighting the importance of T-MCO in copper homeostasis. This research advances our understanding of plant copper tolerance mechanisms and suggests potential applications in crop development. Future work should explore T-MCO's interactions within the copper homeostasis network and its applicability in other plant species.

Dedication

Para Maela, que de haber nacido unas cuántas décadas después estaría haciendo lo mismo.

Contents

List of Tables

List of Figures

Acknowledgements

None of this would have been possible without the help of John, Bob, Miguel, Allie, Ben, Kevin, and Annie, the input from Greg and Sheng Yang, or my friends and family and cats, for somehow keeping alive through all of it.

1. Introduction

Preface:

Understanding the complexities of environmental stress adaptation is a multifaceted endeavor that requires a comprehensive approach. In this study, I will employ Directed Acyclic Graphs (DAGs) to systematically illustrate the inferred causal relationships between variables involved in copper (Cu) tolerance in *Mimulus guttatus*. Figure 1 depicts a funnel-shaped structure, serves as the backbone of this exploration, guiding the reader (and myself) from broad environmental interactions down to specific molecular mechanisms.

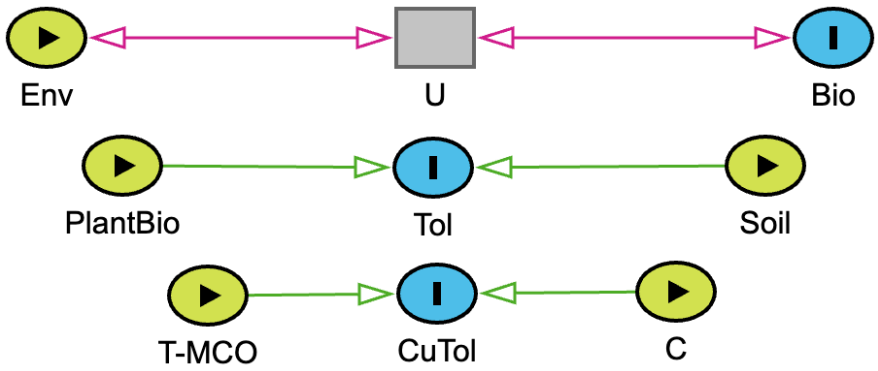

Figure 1. Directed Acyclic Graph (DAG) illustrating the foundational causal relationships in this study. Abbreviations: Env=environment; Bio=living organisms; PlantBio=biological features inherent to plants; T-MCO=MCO in Tol1 locus; Tol=tolerance; CuTol=copper tolerance; C=Cu in the substrate. Nodes represent variables: green triangles for exposure variables, blue bars for outcome variables, and edges represent causal paths (green for direct causal paths, pink for biasing paths). This DAG sets the stage for understanding the complex interactions between environmental and biological factors in copper tolerance. The funnel-shaped structure of the DAG moves from broad concepts (top) to specific details (bottom), narrowing down from general environmental and biological interactions to the specific role of T-MCO in copper tolerance in Mimulus guttatus. The DAG also serves as a preview for the reader of the detailed discussions that will follow in the introduction**.**

DAGs are graphical tools widely used to illustrate causal relationships between variables (Figure 1). The main components of a DAG are simple: nodes (ovals and ocassionaly rectangles)

and edges (arrows). I used open-source software DAGitty[1] to draw all DAGs, so legends will remain consistent throughout: green nodes with a triangle represent exposure variables, blue nodes with bars represent outcome variables, green edges represent causal paths, pink edges represent biasing, non causal paths, and grey indicate "selected" variables that have been explicitly selected for a particular purpose.[1] Pink edges represent cases where the relationship between nodes is not causal or that causality could not be identified.[1,2] This is can happen due to insufficient understanding of the variables, their interaction, their relationship to other unknown variables, the system as a whole, or the *context* in which they exist.[2] For the detailed DAG legend used in www.daggity.net, please see Appendix A.

In this case, the topmost DAG includes one exposure variable (Env for environment), a mediator selected variable (U for unknown or unobserved), and an outcome variable (Bio, meaning living organism). U can represent any type of mechanism involved in stress response, whether adaptive, deleterious, or neutral. The directional of the edges in a DAG shows where causal influence flows (cause -> effect).[1] Here, the dual arrows between variables signifiy multidirectional and/or indeterminate flow of causality.

Biotic and abiotic interactions can take all kinds of shapes and sizes and intensities, and same as the responses they elicit. Aside from the multidirectionality of the edges, emphasis on U being an open-ended or unknown variable is meant to draw attention and represent the inmense range of possibilities that the first DAG could encompass

Building from this basal DAG, the center DAG consists of two exposure variables: PlantBio (biological, intrinsic, internal factors of a plant), Soil (enviromental substrate the plant is in contact with), both pointing towards the middle to the outcome variable Tol (for tolerance, signifying resilient adaptive traits in response to environmental stressors). The direction of causality indicates that the outcome variable is affected by both exposure variables.

Lastly, the most specific DAG consists of two exposure variables: T-MCO (the main candidate gene for confering tolerance to *M. guttatus*) and C (Cu concentration in the mine tailing soil where *M. gutattus* lives, which is beyond tolerance levels for non-locally[3] adapted plants). Both variables point towards the middle to the outcome variable CuTol (referring to *M. guttatus* remarkable Cu resilience).

Figure 1 serves two purposes: to provide the reader with a preview of the detailed discussions that will follow in the introduction, and introduce DAGs and complex systems science, which will be mentioned throughout this document. Starting from simple bivariate linear relationships, DAGs help establish a solid visual and conceptual base for studying multilevel interactions.[2] Using DAGs to build complexity is useful for keeping a record of the process and the logic behind it, so in case something needs to be revised, there is a well-delineated path to retrace. Because the study of adaptation to environmental stressors is inherently complex and multifaceted, I will use these and other tools and concepts commonly used in complex systems science.[2,4–6] I hope this visual summary and this preface helps you understand the flow and interconnectedness of the information that I am trying to convey and that it sets the stage for a deep dive into the molecular and evolutionary mechanisms behind copper tolerance in *Mimulus guttatus.*

1.1 Heavy Metals Never Go Out of Style

Heavy metal pollution is a rapidly growing environmental concern. With the accelerated expansion of industrial activities like mining, agriculture, and chemical manufacturing, there has been a corresponding rise in the levels of pollutants discharged into the environment.[3,7,8] Heavy metals are particularly problematic because, unlike organic pollutants, they do not degrade over time and remain in soils and water indefinitely.[9] Prolonged exposure to heavy metals can lead to

their accumulation within the cells and tissues of living organisms, which can, in turn, facilitate their infiltration into food webs and produce long-lasting health and environmental issues.[9–12]

The role of heavy metals in biological systems, however, is more nuanced and expands beyond their detrimental effects as pollutants. Throughout Earth's history, inorganic nutrient availability has been a key driver of the evolution and diversification of life.[9,12,13] Drastic changes in Earth's chemical composition, such as the oxygenation of the atmosphere and oceans during the Great Oxygenation Event, posed many new metabolic challenges for organisms. Nonetheless, it also brought forth novel metabolic opportunities, such as the ability to use the chemical properties of metals for the biosynthesis and metabolism of macromolecules essential for life.[14] This marked an important stepping stone towards the evolution of multicellular life. [9,13]

Millions of years later, organisms have adapted to accommodate the chemical nature of metals into their systems. Their intrinsic poor solubility and strong binding properties make them less bioavailable, so uptake and acquisition mechanisms must be highly effective. Additionally, because metals tend to have high binding affinities but carry vastly distinct functions, fine-tuned homeostasis mechanisms are needed to avoid competition and interference among different metals. When the levels of a particular metallic ion exceed an organism's tolerance threshold, ions of the excess metal can compete with and replace other metal ions in a binding site. This is called mismetalation and can result in the disruption of a protein's function via incorrect protein folding or oxidative damage, which can further generate a cascade of detrimental effects.[9,12,15,16]

Despite their potential toxicity, many of the biochemical reactions that sustain life on Earth would not be possible without heavy metals, hence why they are a fundamental component of living organisms' essential micronutrient array.[12,17,18] Photosynthesis and its role in biological and environmental processes is probably the best example to illustrate this. Primary producers, like plants, capture energy from the sun and convert it into chemical energy, while also releasing

oxygen into the atmosphere and regulating carbon dioxide levels.[12,19] Essential trace metals like Fe, Mn, and Cu are required for plants (and other photosynthetic organisms) to carry out these reactions. Consequently, this makes plants highly susceptible to changes in the chemical composition of their environment, as they absorb these and many other essential micronutrients directly from the soil.[9,12,17,20]

The role essential trace metals play in biochemical reactions vary depending on the element and reaction. Redox reactions stand out among these because they oversee energetic transactions via nature's energy currency: electrons. In biological systems, copper (Cu), much like iron (Fe), act as redox front liners by catalyzing reactions involving energy transfer, storage, and expenditure[9,12,17]. Although this project focuses on copper, its crosstalk with iron in biological systems is important to mention because of how interdependent they are on each other, and how impactful the effects of imbalance of either of these can be.[14,21] Additionally, even though many of these processes are present throughout all life forms, here I will discuss them through plant/copper-tinted lenses.

1.2 Cu in Plant Biology

Cu, despite being much less abundant when compared to Fe, also has a redox potential adequate for biological systems and is a cofactor of a wide array of enzymes. Cu exists in two oxidation states: the reduced Cu+ (cuprous) and the oxidized Cu2+ (cupric) states and can easily switch between them. [10,17,22] This makes Cu highly versatile, enabling it to participate in a variety of essential physiological functions.

In energy metabolism, Cu participates as a cofactor in electron transport chains within both photosynthesis and respiration by facilitating the oxidation of plastocyanin in chloroplasts and cytochrome c oxidase in mitochondria.[10] Additionally, Cu is involved in antioxidant defense as a cofactor of Cu/Zn superoxide dismutase (SOD), which simultaneously oxidizes and reduces

superoxide radicals ($O_2^{\bullet-}$) into oxygen and hydrogen peroxide. Cu also serves as a cofactor for oxidizing enzymes in the biosynthetic pathways of indispensable metabolites such as chlorophyll and ethylene.

Furthermore, Cu is involved in the uptake of other nutrients via roots, like as iron (Fe) and nitrogen (in legumes)[23]. Finally, Cu contributes to structural integrity through its involvement in cell wall metabolism and lignin synthesis.[23] Cu-dependent enzymes like pectin methylesterases, peroxidases, and laccases modify and strengthen cell walls and polymerize lignin precursors, enhancing structural support and pathogen resistance[12,17,23–26]

1.3 Homeostasis and Lack Thereof

Copper (Cu) in the soil predominantly exists in the Cu^{2+} form, as this is more stable under aerobic and oxidizing conditions.[25] Cu^{2+} ions can be adsorbed onto soil particles, complexed with organic matter, or precipitated as minerals. At the root-soil interface (the rhizosphere), root exudates and microbial activity can reduce Cu^{2+} to Cu^{+}, making it more available for plant uptake. Ferric reductase oxidases (FRO4 and FRO5), for example, is up or downregulated depending on copper levels, facilitating this reduction process.[25,27–29]

Copper transporters in plants are highly specific to the form of copper they transport. High-affinity transporters such as COPT1 and COPT2 are localized to the plasma membrane of root cells and take up Cu^{+} ions to be transported into the cell.[17,24,29] Once inside, they are received by chaperone proteins that ensure safe transport and delivery to various cellular compartments where copper is needed. Chaperone proteins like CCH and CCS also play an important role in binding copper to enzymes, such as Cu/Zn superoxide dismutase (SOD), which is essential for antioxidant defense.[17,25,29]

For long-distance transport, Cu^{+} can be handed off to other transporters like P-type ATPases, such as HMA5, which transports copper into the xylem for distribution throughout the

plant.[29] Excess copper is sequestered in vacuoles by transporters like COPT5 and stored in a non-toxic form, aided by metallothioneins (MTs) and phytochelatins (PCs) that bind and detoxify excess copper.[29]

The transcription factor SPL7 is the main regulator of copper homeostasis, activating the expression of genes involved in copper uptake under copper-deficient conditions, including FRO4, FRO5, COPT1, and COPT2. SPL7 also regulates microRNAs such as miR398, which modulate the expression of copper-utilizing proteins, ensuring optimal copper use during deficiency.[24] Additionally, hormonal signals like ethylene and auxin influence copper homeostasis by modulating the expression of copper transporters and chaperones, balancing copper levels in response to developmental cues and environmental stressors.[21,27,27,29]

Given copper's crucial role in numerous vital physiological processes, any imbalance in Cu levels, whether deficiency or toxicity, can lead to significant effects on the plant. When Cu is deficient, the damage stems from Cu-dependent processes being halted due to its absence.[29] Conversely, when there is an excess of Cu, the damage is two-fold: directly via oxidative stress and indirectly by interfering and interrupting processes carried out by other essential metals, like iron, zinc, or magnesium.[16,29] In some cases, like chlorosis, symptoms may appear similar on both sides of the copper stress spectrum. However, their underlying causes are more nuanced. For example, in Cu deficiency, chlorosis results from impaired chlorophyll synthesis due to the lack of Cu necessary for this process. Conversely, in Cu toxicity, chlorosis arises from oxidative stress that damages chloroplast structures.[29]

As previously mentioned, copper deficiency halts several vital processes because enzymes are not able to function without their respective cofactor. Impaired photosynthesis and respiration can lead to stunted growth.[25,26,29] Disrupted ethylene signaling can interfere with processes like fruit ripening and flower development.[25] Additionally, the plant's antioxidant

defenses, and immune responses are weakened, increasing susceptibility to oxidative stress and pathogens. Poor water and nutrient transport occur due to compromised lignin production and weakened cell wall structure, leading to further symptoms like wilting and malformed leaves.[17]

On the other hand, Cu toxicity overwhelms the plant with oxidative stress, causing widespread cellular damage. High levels of Cu disrupt root function and nutrient uptake, leading to poor growth and overall metabolic dysfunction.[29] The oxidative damage from excessive Cu also severely affects the structural integrity of cells and tissues by binding to membrane phospholipids, structural proteins and carbohydrates, and interfering in their synthesis. This leads to symptoms like distorted leaves and impaired reproductive functions.[30,31] Additionally, the disruption of other nutrient uptake can create a cascade of deficiencies and imbalances. Thus, Cu toxicity is particularly detrimental as it compounds the plant's stress by interfering with multiple physiological pathways.[9,21,29]

1.4 Tolerance Mechanisms: Why Should We Be Looking into Natural Variation?

In Cu-rich soils, excess copper can quickly overwhelm copper homeostasis mechanisms, leading to severe copper stress. To counteract these disruptions, plants have evolved remarkable adaptations depending on their genetic toolkit and evolutionary history. Common strategies include upregulating the synthesis of phytochelatins and metallothioneins, which act as specialized storage facilities for excess copper, ensuring efficient sequestration and preventing damage to the plant’s systems.[17,25,32] Another common strategy is enhanced antioxidant capacity through the increased expression of enzymes like superoxide dismutase (SOD), catalase (CAT), and glutathione reductase (GR), effectively mitigating oxidative stress caused by excess copper.[25] Both methods are found even in species that have other more extreme tolerance mechanisms

because they are an important complement to dealing with current oxidative stress that their other mechanisms can't mitigate.[33,34]

Research on copper tolerance and metal tolerance in general has been dominated by studies on crops and model organisms.[26,26,30,35–37] Anthropogenic activities, such as industrial pollution and intensive agriculture, have negatively impacted soil health. Copper has been widely used as a main ingredient in fungicides for various crops over the years, most notably in vineyards in the form of Bordeaux mixture and in citrus orchards.[7,34,38,39] Its excellent redox properties make it an effective antimicrobial agent, but its accumulation in the soil over time now affects crop yield and water quality. However, the urgent focus on crops and model organisms arises not only due to their direct application but also because of the wealth of genetic and molecular tools available for these systems. Studies on these plants have provided valuable insights into the identification of key genes and pathways involved in metal homeostasis.[14,32,36,37] These controlled environments allow for precise study of genetic and molecular responses to copper stress, leading to practical applications in crop improvement and environmental management.

On the other hand, plants that have been exposed to naturally metal-toxic soils for thousands or millions of years have evolved sophisticated and highly specialized strategies for dealing with stress.[40] Hyperaccumulation happens when plants like *Arabidopsis halleri, Thlaspi caerulescens,* and *Pteris vittata* store high levels of metals, including copper, in their aboveground tissues.[41,42] They localize metals to compartments in leaf or shoot cells or bind them to chelators like organic acids and amino acids, reducing potential toxicity and protecting vital cellular processes.

Conversely, excluders or evaders, employ opposite strategies to avoid copper toxicity. For example, *Silene paradoxa* increases the methylation of carboxyl groups in pectin, reducing

negatively charged copper binding sites and thus minimizing copper accumulation.[43,44] Interestingly, it contrasts well documented strategies seen in multiple plants, like in in citrus. Citruses modify their cell walls to decrease methylation and therefore open up additional binding sites that immobilize copper, preventing it from entering the cellular cytoplasm and interfering with vital processes.[26,45,46]

Studying natural variation is crucial for a comprehensive understanding of Cu as well as more general heavy metal tolerance. As mentioned in section 1.1, the history of nutrient availability significantly impacts how organisms interact with their environment. Species exposed to copper (or other metals) over long periods or in extreme conditions, such as serpentine soils and volcanic soils, have developed unique evolutionary strategies to manage multiple harsh conditions.[40] The imprint of all these adaptive strategies provides a broad genetic pool that contribute to plant resilience in ways such that ancestral variation can give rise to modern-day adaptations.[47,48] Evolutionary wisdom! These resilient plants are essential for the assembly of communities that make-up dynamic ecosystems, as they influence the soil's chemical composition, water dynamics, and create diverse microhabitats that can support a wide range of organisms both above and below ground. They are integral to maintaining biodiversity and ecological resilience in otherwise inhabitable environments.

1.5 Monkeyflower-Covered Mine Tailings

Mine tailings are a great concern for environmental and human health. The abundance of heavy metals in the soil makes it highly toxic and contributes to the oxidation of sulfide minerals, resulting in a drastic acidification of the soil. Lower pH levels make metals more soluble, which can pollute aquifers as well as surface bodies of water.[49–51] The soil structure tends to be fine or sandy with low water holding capacity, leading to water stress and erosion. This can also contribute to polluting nearby soils via wind. Additionally, the very low pH affects the amount of

organic matter in the soil as well as the availability of essential nutrients necessary for biological functions, such as phosphorus, nitrogen, and potassium, thereby making living conditions even harsher for macro and microorganisms.[49]

However, some organisms have managed to locally adapt to these extreme conditions. Mimulus guttatus, the yellow monkey flower, has been studied for decades for its ability to thrive in copper mine tailings. Early studies by Allen and Sheppard (1971), who first described copper-tolerant populations of *M. guttatus* inhabiting copper mine tailings in California. These populations exhibited a remarkable ability to survive and thrive in the presence of high copper concentrations, which are typically toxic to most plants. This initial observation laid the foundation for subsequent investigations into the genetic basis of copper tolerance in *M. guttatus*.

Following these early studies, comprehensive genetic analyses were conducted,[52] including crosses between tolerant and non-tolerant individuals, to elucidate the inheritance pattern of copper tolerance. Their findings suggested that copper tolerance in *M. guttatus* is primarily controlled by a single dominant locus, termed Tol1.[53,54] However, at that time, the precise genomic location and the underlying gene(s) responsible for the tolerance phenotype remained unidentified.

To further dissect the genetic architecture of copper tolerance[54], a high-resolution mapping approach using near-isogenic lines (NILs) was employed. By utilizing recombinant inbred lines derived from crosses between tolerant and non-tolerant parental lines, they were able to narrow down the location of the Tol1 locus to a specific genomic region on linkage group 9. Nonetheless, the repetitive nature of this peri-centromeric region posed significant challenges for the assembly of large genomic scaffolds, hindering the identification of the tolerance gene(s).

To overcome these limitations and pinpoint the candidate genes within the Tol1 region, a combination of fine-mapping, comparative genomics, and transcriptome analysis was used.[53] A

novel genome assembly for a copper-tolerant *M. guttatus* line was developed, which greatly improved the contiguity and resolution of the Tol1 region. By comparing the tolerant and non-tolerant genome assemblies, gene candidate T-MCO (T for Tol1 locus) was identidied.[53,55]

The T-MCO gene (as it will be referred to here) was found to have multiple copies in the tolerant genome compared to the non-tolerant genome, suggesting a potential role in enhancing copper detoxification or sequestration.[53,55] Furthermore, transcriptome analysis revealed significantly higher (12x) expression levels of T-MCO in tolerant plants compared to non-tolerant plants.[53,55]

Studies on other MCOs structure and function in other systems supported the finding of T-MCO as the main candidate gene. Multi-copper oxidases are a family of enzymes that use multiple copper-binding sites to catalyze oxidation reactions.[56–58] Found across the tree of life, these enzymes have highly conserved copper-binding sites typically organized into Type 1 (responsible for electron acceptance), Type 2, and Type 3 (involved in electron transfer to oxygen).[57] This structure creates a pathway that facilitates efficient catalytic processes for various functions. MCOs can exhibit antioxidant properties by managing reactive metal ions, contribute to metal homeostasis by regulating levels of metals like copper and iron, and participate in the breakdown or synthesis of large molecules such as lignin.[14,56–59]

Even though these studies have provided fascinating insights into the genetic basis of copper tolerance in *M. guttatus*, all the presented lines of evidence so far have been indirect. Previous findings have provided ample evidence for correlation,[53,55] but direct evidence for causation has not yet been obtained. For my thesis, I aim to provide direct evidence that the multi-copper oxidase (MCO) genes, found within the Tol1 locus, play a major role in conferring copper tolerance in *Mimulus guttatus*.

2. Investigating the role of MCOs in copper tolerance in *Mimulus guttatus*

2.1 Introduction

Copper (Cu) is a vital micronutrient required for various physiological processes in plants, including photosynthesis, respiration, and oxidative stress response.[29] However, in excess, Cu can become toxic, causing severe oxidative damage and impairing plant growth. Understanding the molecular mechanisms of Cu tolerance is essential for developing strategies to enhance plant resilience in Cu-contaminated environments.[29]

Mimulus guttatus, a species known for its ability to thrive in high-metal environments, serves as an excellent model for studying metal tolerance.[53] Previous research has identified the Tol1 locus and the multi-copper oxidase (T-MCO) genes as protagonists in Cu tolerance. MCOs are enzymes that catalyze the oxidation of various substrates, using Cu as a cofactor. These enzymes play a crucial role in Cu homeostasis by facilitating the safe transport and storage of Cu ions, thereby mitigating their toxic effects.

This chapter delves into the experimental investigations aimed at confirming the role of MCO in conferring copper tolerance in Mimulus *guttatus*.

2.2 Methods

2.2.1 Initial Setup, and Experimental DAGs.

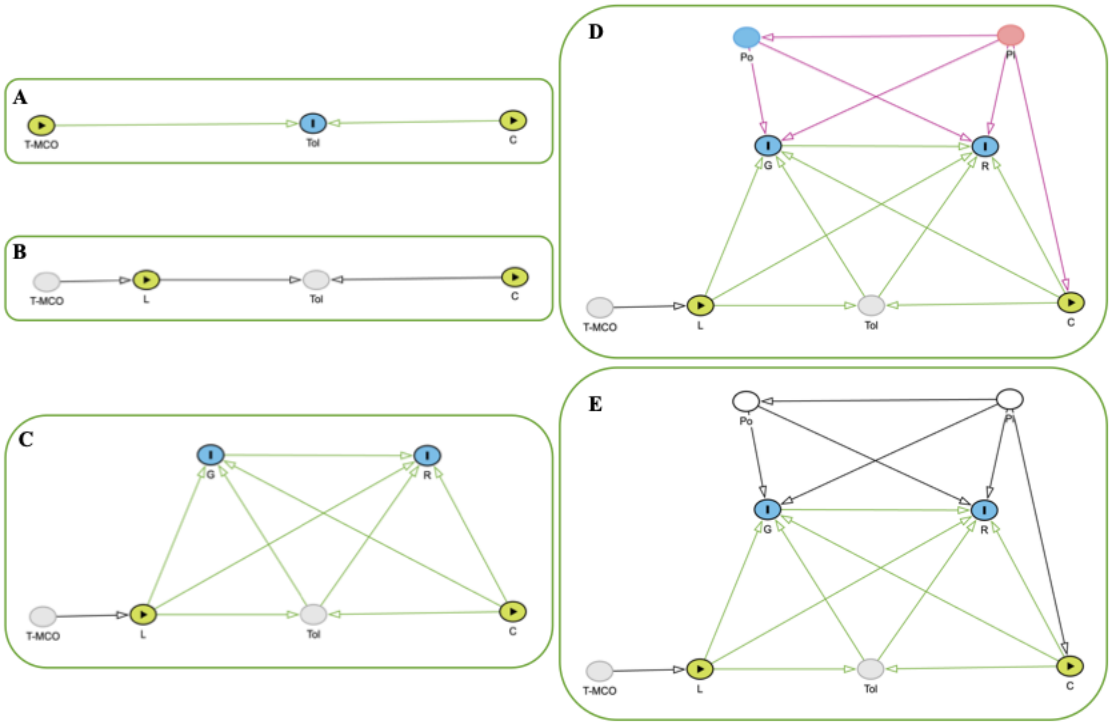

Figure 2: Directed Acyclic Graphs (DAGs) illustrating the experimental design for investigating copper tolerance in *Mimulus guttatus*. Simplified causal relationship between T-MCO gene, tolerance (Tol), and copper concentration (C). (B) Introduction of overexpression lines (L) as a mediator between T-MCO and tolerance. (C) Measurement of tolerance through proxies: root length (R) and germination (G). (D) Full experimental model including potential confounding factors: plate (Pl) and position (Po). (E) Adjusted model accounting for unmeasured factors. Green nodes with triangles represent exposure variables, blue nodes with bars represent outcome variables. Grey nodes indicate unmeasured variables. Arrows depict hypothesized causal relationships, with green arrows showing causal paths and pink arrows indicating potential biasing paths

The central aim of this study is to provide direct evidence of the role T-MCO plays in copper (Cu) resilience in *Mimulus guttatus*. Here, I will expand on the methodology by bridging Figure 1 (bottom DAG) with Figure 2 (2A) through DAGSs, which serve as the backbone for experimental design, data analysis, and interpretation. As explained in Figure 1, the introduction covered the background knowledge and literature review that informed the DAG seen in Figure 2A.

The DAG contains three variables: two exposure variables (T-MCO and copper concentration (C)) and one outcome variable (copper tolerance (Tol)). The path of interest is T-MCO -> Tol, representing the effect of T-MCO on copper tolerance. While this appears to be a simple linear bivariate interaction in the DAG, the relationship between T-MCO and Tol is complex and cannot be directly measured.

To address this, T-MCO was inserted and overexpressed in a non-tolerant background (L), serving as a mediator between T-MCO and Tol. The overexpression lines were created by collaborators Allie Gaudinier and Srinidhi V Holalu in the Blackman Lab at UC Berkeley. Each line was genotyped and PCR-confirmed to ensure successful transformation and then shipped to the Franks Lab at NCSU (North Carolina State University), where I performed the experiments. Since the location of gene insertion is not controlled, each level of L is expected to have a different genotype and potentially a different effect on tolerance.

The lines were selected to have different levels of tolerance to see a gradient of effects. To complement and cross-check this data, gene expression studies are needed to help elucidate whether these differential effects are proportional to expression level. This could indicate that differences are caused directly or indirectly by the insertion of the transgene.

Aside from the different levels of L, I also intervened through the exposure variable copper concentration (C). Since MCO is constitutively expressed[53], I don't expect differences in expression depending solely on changes in C. However, by testing each line across a wide gradient of copper concentrations, can inform when and if tolerance becomes the main influencing factor and when it reaches a capacity where it can no longer handle more copper.

Different lines have different baselines (intercepts) and different strengths of effect (slopes), so we can expect different lines to react differently to each copper concentration.[2] The

DAGs illustrate this: effects on outcome variables will come from C, from L, and from their interaction. For a full discussion on causal relationships between variables, see Appendix A.

Recognizing that effects on outcome variables are not solely due to exposure variables, I considered potential sources of variability other than the ones of interest (Figure 2D). These variables, known as confounder variables (Figure 2C), affect exposure and/or outcome variables and insert variability that, if not accounted for, could interfere with the quality of the outcome estimates.[2] In this study, I included plate and position and incorporated them in the DAG according to their assumed causal effect on other variables (2D).

While error will occur regardless, using tools like DAGs to causally infer the effects confounders could have on the path of interest, it is possible to determine which variables should be adjusted for (or "controlled)".[1,2,8,60] This understanding allows for accounting for confounder effects rather than attempting to eliminate or ignore them, leading to better estimates.[1,2,8,60] Figure 2E shows how previously pink and blue nodes (plate and position) are now white, indicating they've been adjusted for, meaning their variability has been accounted for in the analysis.[2]

2.2.2 Growth Media Preparation and Protocol Development

Seeds were surface sterilized with 70% ethanol, 10% bleach, and distilled water, followed by a cold treatment at 4ºC for two days. Viability was assessed to ensure that the MCO gene insertion did not disrupt essential functions. Lines were tested for viability on sterile agar plates with media found in Appendix B, which also included seed controls for each species.

To prepare the growth media, I followed the recipe in Appendix B, adjusting as needed for each concentration by adding the corresponding volume of copper sulfate ($CuSO_4$) to achieve the desired concentration. To ensure the integrity of the solutions, I prepared all nutrient solutions needed for the media in small batches and stored them away from UV exposure for no more than three months.

We considered using an alternative media that yielded more transparent gels for imaging purposes. However, after several unsuccessful attempts, this medium was deemed unsuitable due to its extremely fast drying time, solidifying within minutes after being taken out of the autoclave.

2.2.3 Determining Copper Concentration Gradient and Growth Chamber Duration Time

To determine the appropriate copper concentration gradient for the *Mimulus guttatus* experiments, I tested several options. Control lines (MAC and CCC9) were evaluated in plates with 0, 100, and 150 μM Cu to establish the upper concentration limit. For *Arabidopsis thaliana,* without a positive control like in *Mimulus* and with the ease of producing more seeds, I tested the copper concentrations directly with the overexpression lines compared to wild-type (WT) plants. Based on previous data from the *Mimulus* experiments, I started testing at 0 and 100 μM Cu for *Arabidopsis*.

2.2.3.1 *Arabidopsis* Experiments

Based on initial testing, I screened *A. thaliana* lines at 20 μM intervals between 0 and 80 μM Cu, as there was no germination at 100 μM Cu. Growth chamber conditions were maintained at 20°C with a 16-hour light/8-hour dark photoperiod for 15 days. This 15-day duration was determined by observing growth plateauing at the highest concentration (80 μM Cu), allowing sufficient differentiation in growth patterns for statistical analysis.

Each concentration had 20 plates, with 4 transgenic lines and 2 wildtype (WT) seeds per plate. The decision to use 2 WT seeds per plate was to avoid the inconsistencies observed with MAC in *Mimulus* and to ensure enough controls for robust statistical analysis. Initially, the first

four transgenic lines were validated in 0 and 100 μM Cu to assess viability. Twenty plates per concentration were chosen to provide sufficient replication for statistical hypothesis testing.

2.2.4 Data Collection

I recorded six main variables: genetic line (Line), copper concentration (Cn), plate identifier (Plate_Cu_Cn), position within the plate (Position), germination (yes/no), and primary root length (R). Germination was assessed visually. For R measurements, I scanned the plates after 15 days and used ImageJ software to trace and measure root lengths.

I included the position within the plate and the plate identifier to account for variability arising from these factors. This would provide more accurate estimates for variables of interest like Line and Cn, and it also helped me test my experimental design by assessing if the variables were producing excessive error, indicating a need for adjustments.

2.2.5 Statistical Analysis

For the analysis, I began with exploratory data analysis (EDA) to understand the distribution and relationships among variables. I conducted separate analyses for germination and primary root length (R) to simplify statistical methods, given the nature of each variable: germination data is binomial (Bernoulli), while R is continuous positive (gamma).

Following the EDA, I used Fisher's exact tests to evaluate the relationships between variables and inform further model fitting for germination within each subset. Based on these results, I fitted a mixed-effect logistic regression (also called generalized linear mixed effect model or GLMM) appropriate for Bernoulli variables. For R, I performed ANOVAs within each concentration subset to understand the effects within concentrations.

For the statistical analysis, I used R (R Core Team, 2021) as my primary software environment. I used the following packages: `rethinking` for Bayesian multilevel modeling[2],

`lme4` for fitting generalized linear mixed-effects models (Bates et al., 2015)[61], `ggplot2` for data visualization, `dplyr` for data manipulation within the Tidyverse suite,[62] and `dagitty` to draw DAGs.[1]

2.3 Results

2.3.1 Establishing concentration gradients

In the initial screening, *Mimulus guttatus* lines were tested at various copper (Cu) concentrations to determine tolerance limits. At 150 μM Cu, the positive control CCC9 exhibited reduced growth but was able to produce roots, while the non-tolerant MAC showed negligible growth. Based on these results, the upper limit for subsequent experiments was set at 100 μM Cu. For Arabidopsis thaliana negative control (WT), seeds did not germinate at 100 μM Cu, indicating severe toxicity. Therefore, the concentration gradient for Arabidopsis was set at 20 μM intervals up to 80 μM Cu. Seed placement for the *Arabidopsis* be seen in Figure 4.

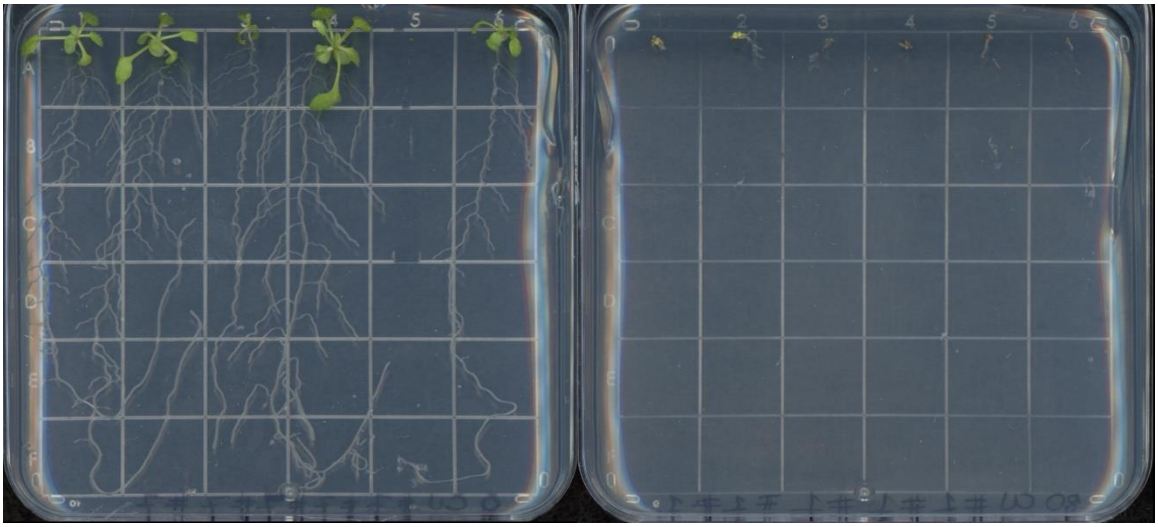

Figure 3: Comparison of *Arabidopsis thaliana* root growth under different copper concentrations after 15 days. Left: Plate 7_0 (0 μM Cu) showing lines from left to right: WT, Line 1, Line 2, Line 4, WT, Line 3. Right: Plate 1_80 (80 μM Cu) showing lines from left to right: Line 1, Line 2, WT, Line 3, Line 4, WT. The images illustrate the dramatic effect of copper

concentration on root growth, with 0 μM Cu resulting in overgrown roots that may lead to measurement inaccuracies. This suggests that future experiments should consider shorter growth periods, as variations between lines at 80 μM Cu are minimal after 15 days.

2.3.2 Experimental setbacks and adjustments

Several setbacks occurred with the *M. guttatus* lines. Having established lower and upper concentrations with the control lines, I started screening overexpression lines in both 0 and 100 μM Cu. Initially, I planted two rows of five seeds, with one of each control seed per row for reference (Figure 5). CCC9 roots grew excessively, crowding the plate, so I removed them for the screening stage to prevent interference with the lines being tested. Instead, I added another MAC seed to each row because its germination was inconsistent, often resulting in one or no germinations per plate.

I selected lines that showed consistent germination in 0 μM Cu plates and set aside the rest. In plates with 100 μM Cu, transgenic lines germinated and grew roots but exhibited a phenotype we referred to as "octopus roots" (Figure 5). As the name suggests, the roots resembled octopus tentacles: primary root growth was stunted and replaced by increased lateral branching. Additionally, the roots would grow out of the media and into the empty space between the media and the plate lid, and the tips looked stunted, sometimes with burnt tips and/or with purple pigmentation.

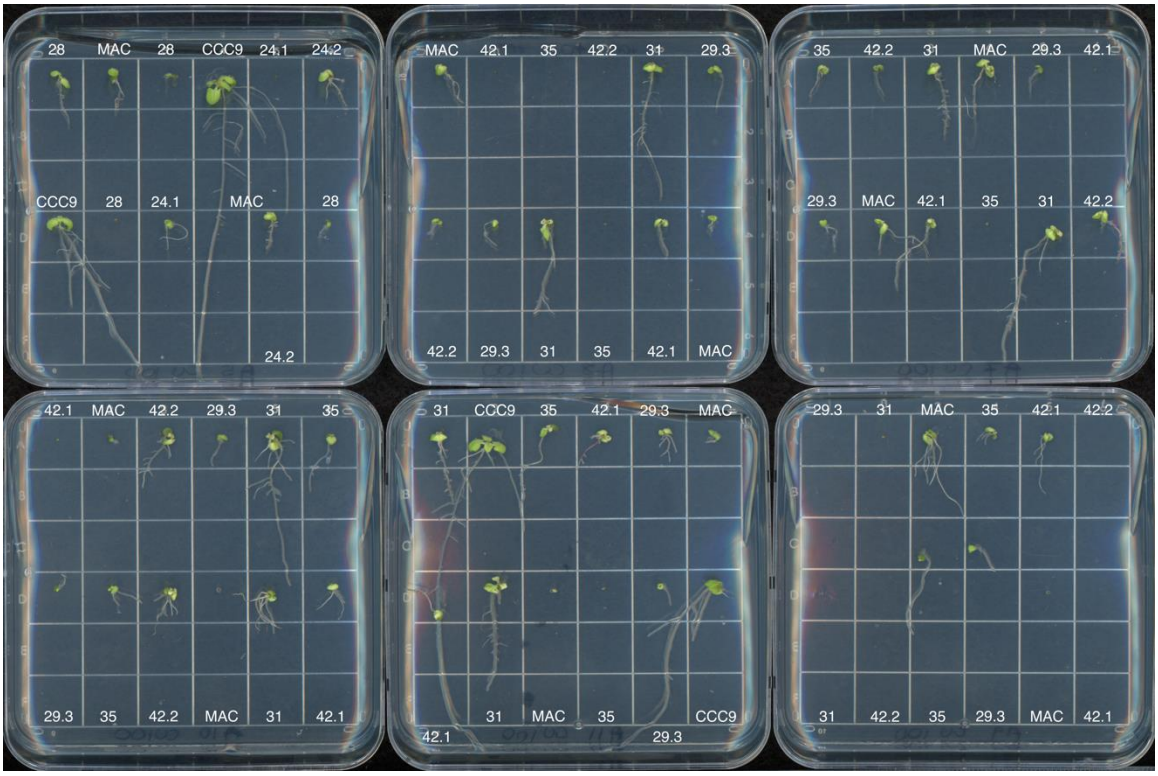

Figure 4: *Mimulus guttatus* seedling growth on plates with 100 µM Cu concentration. The images showcase the 'octopus roots' phenotype observed in most overexpression lines, characterized by increased lateral branching, stunted primary root growth, and roots growing out of the media. The phenotype is less pronounced in the naturally tolerant positive control CCC9 and in better-performing lines like 31. Notable pigment changes are visible in line 42.1 (center bottom plate, bottom row). These observations suggest that future experiments should consider hydroponic or soil-based methods to maintain homogeneous root conditions and allow for more detailed analysis of root phenotypes.

After consulting with the larger copper group in the Franks-Blackman-Willis labs, I tried several potential solutions. First, I adjusted the angle of the plates in the growth chamber to a slight backward tilt. Using an acrylic stand I ensured all plates had the same tilt and rows were equidistant to prevent shadowing from the light source. Unfortunately, this did not work.

Next, I filled the plates to the top with media, let them dry, sealed them, and made small holes at the top to insert the seeds. This was also unsuccessful. When turned upright, even though the media was dry, it would migrate downwards and therefore lower the level at which the seeds were placed away from the top holes. This made the imaging for root tracing and the germination

assessment very difficult. The roots did not show much growth either, likely due to insufficient air circulation.

Subsequent attempts would involve placing a thin mesh on top of the gel to keep the roots in the media, trying a different type of plate, and exploring hydroponic or soil-based growth methods. Nonetheless, due to time constraints and the need for preliminary data to fulfill program requirements, I had to temporarily shift my efforts to the *A. thaliana* lines instead.

2.3.3 *Arabidopsis* lines

2.3.3.1 Experimental adjustments

Based on initial control testing, I screened *Arabidopsis thaliana* (WT) and transgenic lines at 0 and 100 μM $CuSO_4$. While WT plants appeared healthy at 0 μM, there was no germination at 100 μM due to *Arabidopsis's* lower copper tolerance. At 80 μM, I observed some germination and root growth, so I established a concentration gradient of 20 μM intervals between 0 and 80 μM.

The growth chamber conditions were maintained at 20°C with a 16-hour light/8-hour dark photoperiod for 15 days. Each concentration had 20 plates, with 4 transgenic lines and 2 WT seeds per plate (Figure 4). Seeds were sown on sterile agar plates containing the custom media, adjusted to the appropriate concentration for each treatment group. The 15-day duration was based on observing growth plateauing at the highest copper concentration (80 μM Cu), allowing sufficient differentiation in growth patterns for preliminary statistical analysis. The first four lines tested demonstrated varying levels of copper tolerance, making them suitable for subsequent gene expression experiments.

2.3.3.2 Data Analysis

The exploratory data analysis provided insights into the distribution and variability of primary root lengths (R) across different copper concentrations and genetic lines. At 0 μM Cu, primary root lengths were generally longer, indicating healthy growth. As copper concentration increased to 20, 40, 60, and 80 μM, the distribution shifted towards shorter roots, indicating increased copper toxicity.

For germination, I observed that the proportion of germinated seeds varied oddly across different copper concentrations and genetic lines. Surprisingly, at 0 μM Cu, the WT line had the lowest germination rate (Figure 6). As the copper concentration increased to 20 and 40 μM, germination rates improved, but then declined again at 60 and 80 μM Cu. Lines 1 and 2 exhibited consistently high germination proportions across all copper concentrations. All Line 1 seeds germinated across all concentrations, and Line 2 seeds also had perfect germination in concentrations 20 and 40. Line 3 and Line 4 showed a decline in germination proportions at higher concentrations, but less severe than WT.

To analyze root length data, I separated it by copper concentration subsets and performed one-way ANOVAs (Table 2) for each subset to better understand the combined effect of line and copper concentration on germination and root length (R) within each concentration. The results showed, as expected, that the effect of genetic line on R was not significant at 0 μM Cu but became significant at higher copper concentration. The effect peaked at 40 μM and then started to slowly decline at 60 μM, and more rapidly decline at 80 μM (Figure 7). So, the results of the ANOVA support our assumption that the overexpression of T-MCO does affect the plants' ability to tolerate higher Cu concentrations than it would in the absence of the gene. In the absence of Cu treatment, 0 μM, there is no significant Line effect because the plants are not under Cu stress and therefore it doesn't matter if you have or don't have T-MCO.

The germination data did not respond as excepted at all. At 0 μM, WT showed the lowest number of germinated seeds across all lines and all concentration, which is the opposite of what I expected given that WT received no intervention and should have had similar numbers of germinated seeds as the rest. WT seeds germinated less in 0 μM than they did 80μM. This was the first indicator of an experimental error.

Due to Line 1 having zero zeroes and Line 2 having only a small amount (germination being a Bernoulli yes or no 1 or 0 variable), I used Fisher's exact test (Table 1) instead of the chi-squared test to analyze germination proportions, providing a more accurate analysis given the small sample sizes and sparse data. The effect of line appeared significant compared to wild type in concentrations 0, 60, and 80 μM Cu, which has highly unexpected for me.

To get a more nuanced idea of how variables were interacting and get closer look at how each individual line performed at each Cu concentration, I fitted generalized linear mixed models (GLMMs) to the subsets. All relevant statistical details, including sample sizes and the results of the ANOVA, Fisher's exact test, and GLMM analyses, are presented in Tables 1, 2, 3, and 4. In germination models. Consistent with Fisher's test (Table 1) Lines 1 and 2 consistently showed high germination success across all copper concentrations, while the wild type (WT) displayed the lowest germination rate at 0 μM Cu (Figure 5, Table 3). Germination success improved at intermediate copper concentrations (20 and 40 μM) but declined again at higher concentrations (60 and 80 μM) (Table 3). The inclusion of random effects for plate and position in the analysis allowed for a more accurate estimation of these effects, accounting for variability introduced by different experimental conditions. The results indicated that genetic Line significantly impacted germination rates, especially at the extreme copper concentrations of 0, 60, and 80 μM (Table 3).

For root length, significant differences were observed among the genetic lines at all copper concentrations except 0 μM. At 20 μM Cu, Lines 1 and 2 had significantly longer roots

compared to WT (Table 4). Line 2 demonstrated the longest roots across most concentrations, while Line 1, despite its high germination rates, showed less robust root growth compared to other transgenic lines. The analysis also highlighted that plate-to-plate variation generally increased with copper concentration for both germination and root length, while position effects remained relatively stable. The residual variance in root length models increased with copper concentration, suggesting greater individual variability under higher stress conditions (Table 4)

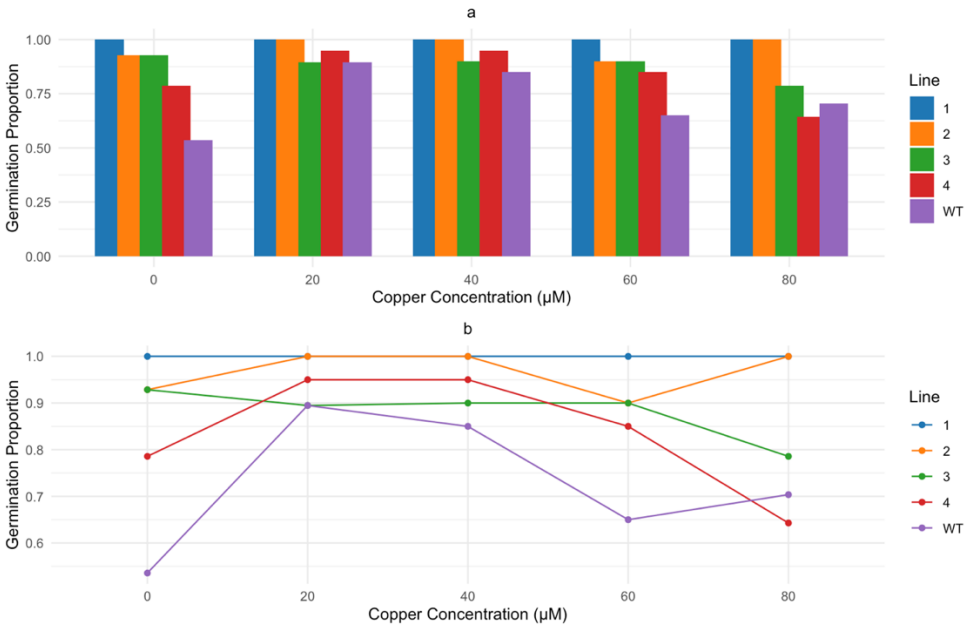

Figure 5: Proportion of germinated *A. thaliana* seeds across experimental Cu concentrations. A) Bar plots show the proportion of seeds germinated per line at each concentration subset: 0, 20, 40, 60, and 80 µM. B) Line graph depicting germination proportions across the same Cu concentrations for each genetic line. The bar graphs provide a clear comparison of germination proportions at each copper concentration, while the line graph illustrates the trend across concentrations for each genetic line. The data indicate that germination proportions tend to decrease as copper concentration increases. Notably, Lines 1 and 2 exhibit nearly 100% germination at lower concentrations (0-40 µM), significantly higher than WT. The anomaly of Line 1 and 2's high germination proportions is reflected in the large error values observed in the GLMM analysis. Note: line and bar plots represent direct proportions, not averaged values

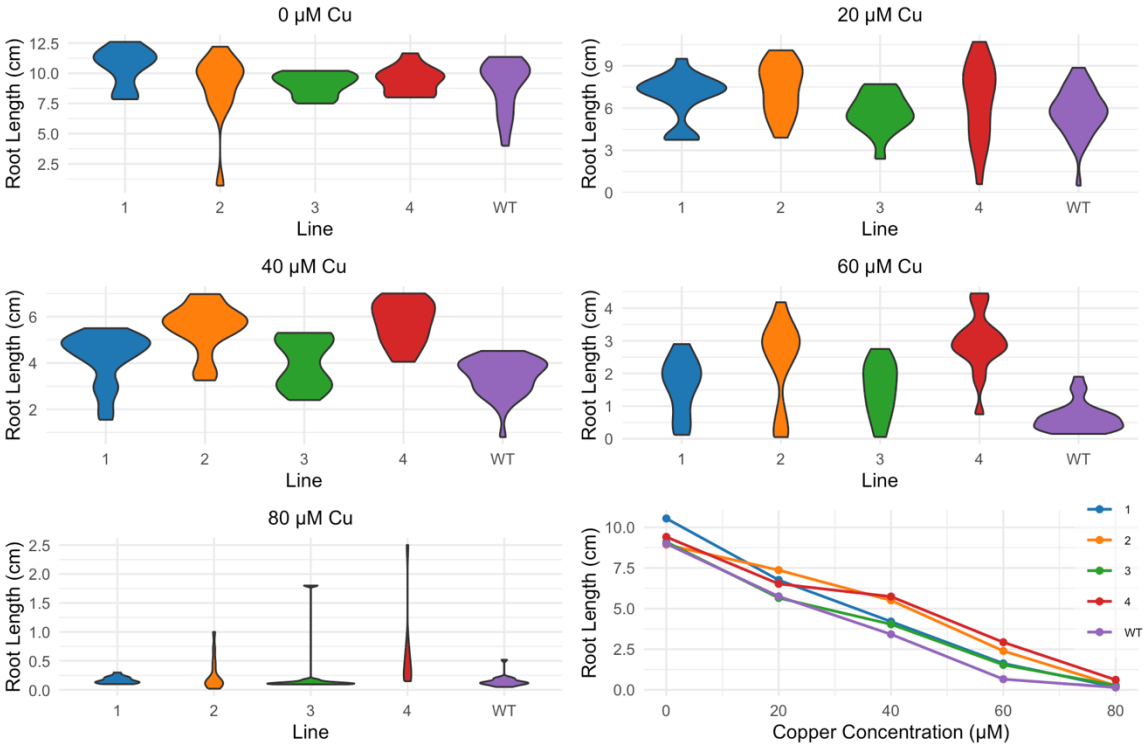

Figure 6: Primary root length analysis *Arabidopsis thaliana* under copper stress. Violin plots illustrate the distribution of root lengths at 0, 20, 40, 60, and 80 µM Cu concentrations. The bottom right panel presents average root lengths for each genetic line across the copper concentration gradient. Results show a consistent decrease in root lengths with increasing copper concentrations. The wild type (WT) exhibits the most severe reduction, particularly at higher concentrations. Line 2 maintains the longest root lengths, indicating the highest copper tolerance, while Lines 1 and 3 show moderate tolerance. Line 4 displays a decline like WT. These observations are further supported by ANOVA and GLMM results

Table 1: Fisher's Exact Test Results for Germination in *A. thaliana* Across Experimental Copper Concentration Subsets

Cu Cn	P
0	0.00120938***
20	0.33470911

40	0.16632503
60	0.00623964**
80	0.01648892**

Note: *p≤0.05; **p≤0.01; ***p≤0.001

Table 2. ANOVA Summary for Primary Root Length in *Arabidopsis thaliana* Across Experimental Copper Concentration Subsets

Cu Cn	df	Sum Sq	Mean Sq	F value	Pr(>F)
0	4	24.92	6.229	1.723	0.156
20	4	45	11.253	3.162	0.017 **
40	4	94.33	23.582	25.13	1.2e-14 ***
60	4	62.72	15.681	21.21	1.88e-12 ***
80	4	1.505	0.3763	2.779	0.0346 **

Note:

Formula: Roots~Line

*p≤0.05; **p≤0.01; ***p≤0.001

Table 3: GLMM Summary for Germination in *Arabidopsis thaliana* Across Experimental Copper Concentration Subsets

Cu Cn	
0 µM	**Fixed Effects**

	Parameter	**Estimate**	**SE**	**Z**	**P**
	Intercept	0.1003	0.4705	0.213	0.8312
	Line 1	20.4642	7511.789	0.003	0.9978
	Line 2	2.7774	1.1932	2.328	0.0199**
	Line 3	2.8504	1.1964	2.383	0.0172**
	Line4	1.3301	0.7887	1.686	0.0917
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.322	0.5675		
	Position	0.292	0.5404		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 84 Plates: 14 Position: 6	-33.6	67.1	77	
20 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**Z**	**P**
	Intercept	2.199	0.5469	4.022	0.0001***
	Line 1	19.29	10320	0.002	0.999
	Line 2	19.03	9061	0.002	0.998
	Line 3	-0.0004	0.9352	0	1
	Line4	0.8101	1.184	0.684	0.494

	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.1474	0.3839		
	Position	0	0		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 117 Plates: 20 Position: 6	-23.1	46.3	110	
40 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**Z**	**P**
	Intercept	1.7346	0.4428	3.917	0.0001***
	Line 1	19.61	9626	0.002	0.998
	Line 2	19.55	9380	0.002	0.998
	Line 3	0.4626	0.867	0.534	0.594
	Line 4	1.2098	1.1175	1.083	0.279
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0	0		
	Position	0	0		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	

	Total: 120	-27.4	54.8	113	
	Plates: 20				
	Position: 6				
60 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**Z**	**P**
	Intercept	0.619	0.3315	1.867	0.0618
	Line 1	19.56	591.21	0.033	0.9736
	Line 2	1.5782	0.8157	1.935	0.053*
	Line 3	1.5782	0.8157	1.935	0.053*
	Line 4	1.1156	0.7086	1.574	0.1154
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0	0		
	Position	0	0		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 120				
	Plates: 20	-47.4	94.7	113	
	Position: 6				
80 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**Z**	**P**
	Intercept	0.9923	0.5028	1.974	0.0484**

	Line 1	19.63	7730.04	0.003	0.998
	Line 2	19.73	7640.35	0.003	0.998
	Line 3	0.5241	0.8189	0.64	0.522
	Line 4	-0.3006	0.7281	-0.413	0.6797
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.8033	0.8963		
	Position	0	0		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 82	-32.1	64.3	75	
	Plates: 14				
	Position: 6				

Notes:

*p≤0.05; **p≤0.01; ***p≤0.001
Formula: Germination ~ Line + (1 | Plate) + (1 | Position)
Family: Binomial
Link function: logit

Table 4: GLMM Summary for Root Length Across Experimental Copper Concentration Subsets

Cu Cn					
0 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**t**	**P**

	Intercept	2.200507	0.07599	28.958	<2e-16 ***
	Line 1	0.161054	0.104198	1.546	0.122
	Line 2	-0.018547	0.113767	-0.163	0.87
	Line 3	0.003572	0.109934	0.032	0.974
	Line 4	0.044968	0.111948	0.402	0.688
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0	0		
	Position	0.0005224	0.02286		
	Residual	0.0395314	0.19883		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 66	-156.6	313.3	58	
	Plates: 14				
	Positions: 6				
20 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**t**	**P**
	Intercept	1.709169	0.079564	21.482	<2e-16 ***
	Line1	0.187185	0.087576	2.137	0.03256 *
	Line2	0.268821	0.085872	3.13	0.00175 **
	Line3	0.008966	0.091415	0.098	0.92187
	Line4	0.124053	0.088145	1.407	0.15932

	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.01703	0.1305		
	Position	0	0		
	Residual	0.06485	0.2546		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 109	-230.3	460.5	101	
	Plates: 20				
	Positions: 6				
40 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**t**	**P**
	Intercept	1.23251	0.052	23.701	<2e-16 ***
	Line1	0.19174	0.06726	2.851	0.00436 **
	Line2	0.4675	0.06736	6.941	3.90e-12 ***
	Line3	0.1505	0.07129	2.111	0.03475 *
	Line4	0.50434	0.06865	7.346	2.03e-13 ***
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.0049538	0.07038		
	Position	0.0003532	0.01879		
	Residual	0.0462888	0.21515		

	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 111	-163.2	326.4	103	
	Plates: 20				
	Positions: 6				
60 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**t**	**P**
	Intercept	-0.4889	0.1629	-3	0.0027 **
	Line1	0.946	0.1989	4.756	1.97e-06 ***
	Line2	1.2925	0.2038	6.341	2.29e-10 ***
	Line3	0.9313	0.2179	4.275	1.92e-05 ***
	Line4	1.5147	0.2003	7.564	3.91e-14 ***
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.025087	0.15839		
	Position	0.009481	0.09737		
	Residual	0.265852	0.51561		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 98	-123.2	246.5	90	
	Plates: 20				
	Positions: 6				

80 µM	**Fixed Effects**				
	Parameter	**Estimate**	**SE**	**t**	**P**
	Intercept	-1.875074	0.002138	-877.16	<2e-16 ***
	Line1	0.147138	0.00216	68.13	<2e-16 ***
	Line2	0.390762	0.002199	177.73	<2e-16 ***
	Line3	0.140582	0.002112	66.58	<2e-16 ***
	Line4	1.223795	0.002223	550.49	<2e-16 ***
	Random Effects				
	Group	**Var**	**Std.Dev.**		
	Plate_Cn	0.252525	0.50252		
	Position	0.003009	0.05486		
	Residual	0.495289	0.70377		
	Model Info				
	Obs	**logLik**	**Deviance**	**df.resid**	
	Total: 66	43.6	-87.2	58	
	Plates: 14				
	Positions: 6				

Notes:

*p≤0.05; **p≤0.01; ***p≤0.001
Formula: Roots ~ Line + (1 | Plate) + (1 | Position)
Family: Gamma
Link function: log

2.4 Discussion

The central aim of this study was to validate the role of multi-copper oxidase within *Mimulus guttatus* Tol1 locus (T-MCO) in conferring copper tolerance to the plant. Previous studies have provided ample indirect evidence suggesting the involvement of T-MCO within the Tol1 locus, but further scientific advances were needed to isolate its effects and confirm its role.[54,54] Collaborators at UC Berkeley successfully created lines overexpressing T-MCO in a non-tolerant genetic background of *Mimulus guttatus* and *Arabidopsis thaliana*, which I used to experimentally observe and validate its effects on the plant's copper resistance. My results confirm that the presence of this gene greatly enhanced copper resilience in *Arabidopsis* overexpression lines, as they significantly outperformed wild-type plants in both germination and primary root growth.

Studies have established the role of MCOs in copper homeostasis, primarily in bacteria and fungi, where these enzymes are central to copper detoxification processes.[56,57] Nevertheless, in plants, even though MCOs are known to participate in other important biological processes, such as lignin synthesis, MCO-centered copper tolerance mechanisms had not yet been observed. This novel mechanism is an exciting contribution to our understanding of plants' genetic toolkits for coping with copper toxicity in copper-rich soils. This is particularly relevant given the increasing levels of copper contamination due to the mismanagement of industrial waste and the overuse of copper-based fungicides.

Here I report indications of the same effect in non-tolerant *Mimulus guttatus* backgrounds; however, more work is needed to fully confirm and understand the function of MCO in *Mimulus guttatus*. Preliminary observations suggest that naturally tolerant *Mimulus guttatus* lines, such as CCC9, exhibit an even stronger tolerance effect compared to transgenic lines overexpressing T-MCO alone. These findings further support the notion that additional

genes or gene interactions are involved in the copper tolerance system. Future studies are needed to determine whether these tolerance mechanisms work in tandem indirectly or if they are actively and directly interacting within the same subsystem. Obtaining a more complete spectrum of genetic interactions and synergistic molecular mechanisms is necessary in understanding how *Mimulus guttatus* resilient profile was shaped and how it provides such robust copper tolerance.

Additionally, in an evolutionary context, the discovery of this novel MCO-centered copper tolerance mechanism enhances the value of *Mimulus guttatus* as a model system for studying rapid adaptation to extreme environments. Remarkable advances have been made—and are in the process of being made—in understanding how other *Mimulus* populations have adapted to serpentine soils, which are notoriously challenging environments due to their high metal content and low nutrient availability.[63] Looking ahead, comparing both evolutionary and molecular mechanisms could provide a powerful approach to elucidating how these adaptive traits evolve.

2.4.1 Lessons Learned and Areas for Improvement

Overall, the lessons I learned during this study follow a similar theme: the importance of carefully considering the variables at hand, particularly in experimental design and scientific inquiry. This became evident during the *Mimulus* root growth assays, where I observed the "octopus root" phenotype—a root structure characterized by its stunted primary root growth and increased branching that deviates from typical growth patterns. I was determined to make the seedlings grow in a linear fashion, like *Arabidopsis* seedlings, to achieve a cleaner read on primary root length. However, in retrospect, I realize that I was trying to study these responses with a 2D perspective when a 3D approach was clearly needed. This oversight eventually led to my missing the invaluable opportunity to use the *Mimulus* overexpression lines to their full potential.

Instead, I should have considered the greater context of this stress response and focused on optimizing and characterizing it in a way that would allow for quantitative analysis. For example, soil and hydroponic assays likely would not have encountered the same problems as plate assays and could have easily allowed for the measurement of mass in grams instead of just length. This approach would have provided a far better picture of changes in total root growth rather than only accounting for primary root growth. Additionally, I could have taken advantage of the root growth patterns, phenotype, and architecture, characterizing more subtle signs of copper stress, such as the presence of pigments (potential copper chelators), the number of total root shoots compared to burnt tip root shoots, growth rate, and more.

This would still have allowed me to quantify root growth as a variable, but it would also have provided additional information to help characterize the phenotype more thoroughly. Considering this was the first time these experiments were done with these *Mimulus* transgenics, having that thorough phenotype characterization would have been a great resource for further studies. Moreover, I could have used other tools or experimental setups to measure and account for how these variables change over time, potentially revealing very interesting associations by not limiting the observations to a single snapshot in time. This approach would have made it easier to integrate the variables of germination and root growth, for example, which are undeniably correlated because one directly affects the other.

Furthermore, if time had been used as an X-axis in some of the data collection, I could have measured changes or disruptions in vital processes involving copper, such as flowering, shoot and leaf development, or seed viability. Spatial analysis of root growth would have been an interesting way of exploring why these root growth patterns have evolved and how or if root architecture might be directly involved in the copper tolerance mechanism of *Mimulus*. Conducting these experiments in *Arabidopsis* as well could have been very informative,

especially given that most gene descriptions have been done in *Arabidopsis*, providing a stronger reference for many of these parameters. However, characterizing the phenotypes enough to compare primarily against *Mimulus* itself would have been ideal and would have led to more interpretable results.

Another example of the lesson learned came from the seed germination experiments. I overlooked important variables, such as seed age and seed handling (like sterilization procedures), which significantly impacted the germination rates and made the results difficult to interpret. This was a mistake in experimental design, as I failed to recognize how these factors could influence the outcome.

In future studies, I would prioritize using seeds of a similar age, ideally freshly harvested, to ensure consistency across experiments. Additionally, I would carefully document details like the seed batch, age, and any sterilization processes, like how I tracked plate and position effects. This would allow these variables to be accounted for simultaneously with other factors during the analysis and interpretation of the data, reducing variability and improving the reliability of the results. Moreover, having this information documented would serve as reference for future studies, as it would provide a clearer understanding of how these factors influence germination and overall plant development. After all, science is cumulative: every contribution adds to the growing body of knowledge.

And finally, the most important iteration of the lesson learned was the significance of obtaining as comprehensive a picture of the system at hand as possible, particularly when dealing with complex traits like adaptation to environmental stress. Tools like DAGs and causal inference provide a robust framework to build networks of causality that can inform all aspects of scientific research. By understanding the dynamism of our study systems—both within them and in relation to their broader context—rather than viewing them as isolated or individual entities, we can

achieve more interpretable and applicable results that lead to more meaningful science. Embracing and working with experimental error, rather than against it or avoiding it, can be highly informative and provide a more nuanced understanding of our work and its impact.

There is great value in acknowledging the context of our scientific questions beyond the literature review for a grant proposal or scientific paper. Integrative, holistic approaches are necessary in both basic and applied sciences to innovate and make progress in an ethical and inclusive manner. Most of the work we do involves complex systems that are inherently non-linear and rarely exist outside of their context. Furthermore, creating, supporting, and maintaining inclusive, strength-based research and work environments that welcome and support non-linear thinking has value far beyond "just" social dynamics. Because, just like most of our study systems, scientists are also people, and we cannot exist outside of our context.

3. Conclusion

This study set out to validate the role of multi-copper oxidase (MCO) within *Mimulus guttatus* Tol1 locus (T-MCO) in conferring copper tolerance to the plant. Through a series of experiments involving the overexpression of T-MCO in non-tolerant genetic backgrounds of *Mimulus guttatus* and *Arabidopsis thaliana*, the findings confirmed that T-MCO significantly enhances copper resilience. This was evident through improved germination rates and primary root growth under copper stress, highlighting the critical role of this gene in the plant's ability to withstand high copper concentrations.

These results provide strong evidence for the role of MCO genes in copper tolerance. MCO-centered tolerance mechanisms, traditionally observed in bacteria and fungi, are now being recognized in plants, adding a new dimension to our understanding of how plants like *Mimulus guttatus* adapt to extreme environmental conditions, particularly in copper-rich soils.

The findings support and align with previous studies indicating that copper tolerance is not solely dependent on T-MCO; rather, additional genes or gene interactions likely work in tandem to confer this resilience. Observations of stronger tolerance in naturally tolerant lines, such as CCC9, compared to transgenic lines overexpressing T-MCO alone, suggest the presence of complementary genetic factors that warrant further investigation.

The challenges encountered during this research, particularly the need for a more holistic approach in experimental design and the importance of accounting for variables like seed age and root growth dynamics, have provided valuable lessons for future studies. Moving forward, embracing tools like DAGs and causal inference frameworks will be essential in building a more comprehensive understanding of these systems, leading to more reliable and meaningful scientific outcomes.

In conclusion, this dissertation not only validates the role of MCO in copper tolerance but also opens new directions for exploring the genetic and molecular mechanisms underlying this trait. The insights gained here have broader implications for understanding plant adaptation to environmental stressors and could inform strategies for developing crops with enhanced tolerance to heavy metals, contributing to more sustainable agricultural practices in contaminated environments.

Appendix A: DAGs

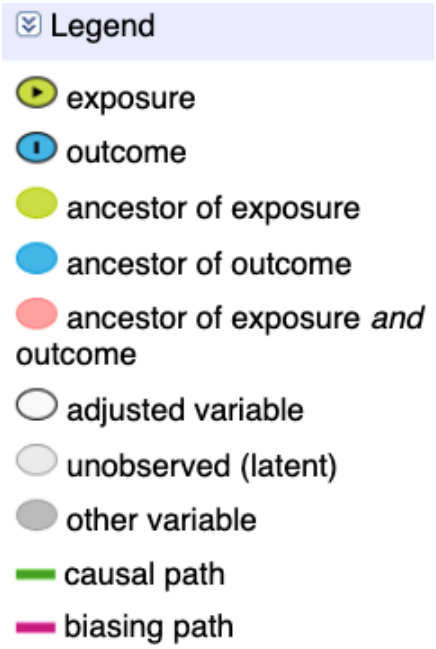

Figure A1: Legend for DAGs using DAGitty, from their online software version http:// dagitty.net.[1]

Appendix B: Media

Copper Medium Recipe (for 1 liter)

Ingredients/Components:

- KNO_3 (1 M): 10 mL
- $MgSO_4$ (1 M): 4 mL
- KH_2PO_4 (1 M): 2 mL
- FeEDTA solution: 2 mL*
- Micronutrient stock: 2 mL*
- Sucrose: 10 g
- $CaCl_2$ (1 M): 1 mL
- KCl (1 M): 10 mL
- MES (FW=195.2): 0.146 g
- Distilled Water: Up to 1 liter

Preparation Steps:

1. Begin with 700 mL of 0.2-micron filtered water on a stir plate.
2. Gradually add each component to the water while stirring.
3. Adjust the pH to 5.7 using 2 M KOH.
4. Add 0.2-micron filtered water to bring the total volume to 1 liter.
5. For solid media, add 7.5 g of agar.
6. Sterilize the medium by autoclaving at 121°C for 15 minutes.
7. For copper concentration adjustments, add the corresponding volume of $CuSO_4$ solution to achieve the desired concentration.

***Special Instructions:**

- *FeEDTA Preparation*: Dissolve 7.45 g Na_2EDTA in 1 L water, then add 5.57 g $FeSO_4 \cdot 7H_2O$.

- **Micronutrient Stock Preparation** (for 1 liter):
    - 2.86 g H_2BO_3
    - 1.81 g $MnCl_2 \cdot 4H_2O$
    - 0.11 g $ZnCl_2$
    - 0.05 g $CuCl_2 \cdot 2H_2O$
    - 0.025 g $Na_2MoO_4 \cdot 2H_2O$

Usage:

This recipe is for plates with a copper concentration of 0. To increase the copper concentration, use a 200 mM Cu^{2+} solution ($CuSO_4$). For a 100 μM Cu^{2+} concentration, add 20 μL of the copper solution to every 40 mL of media.

References

1. Textor, J., Van Der Zander, B., Gilthorpe, M. S., Liśkiewicz, M. & Ellison, G. T. H. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int. J. Epidemiol.* dyw341 (2017) doi:10.1093/ije/dyw341.

2. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. (Chapman and Hall/CRC, New York, 2016). doi:10.1201/9781315372495.

3. Adrees, M. *et al.* The effect of excess copper on growth and physiology of important food crops: a review. *Environ. Sci. Pollut. Res.* 22, 8148–8162 (2015).

4. Saavedra, S., Bartomeus, I., Godoy, O., Rohr, R. P. & Zu, P. Towards a system-level causative knowledge of pollinator communities. *Philos. Trans. R. Soc. B Biol. Sci.* 377, 20210159 (2022).

5. Battiston, F. *et al.* The physics of higher-order interactions in complex systems. *Nat. Phys.* 17, 1093–1098 (2021).

6. Reed, P. M. *et al.* Multisector Dynamics: Advancing the Science of Complex Adaptive Human-Earth Systems. *Earths Future* 10, e2021EF002621 (2022).

7. Ballabio, C. *et al.* Copper distribution in European topsoils: An assessment based on LUCAS soil survey. *Sci. Total Environ.* 636, 282–298 (2018).

8. Liu, Y. *et al.* Causal inference between bioavailability of heavy metals and environmental factors in a large-scale region. *Environ. Pollut.* 226, 370–378 (2017).

9. Krämer, U. Metal Homeostasis in Land Plants: A Perpetual Balancing Act Beyond the Fulfilment of Metalloproteome Cofactor Demands. *Annu. Rev. Plant Biol.* 75, annurev-arplant-070623-105324 (2024).

10. Mir, A. R., Pichtel, J. & Hayat, S. Copper: uptake, toxicity and tolerance in plants and management of Cu-contaminated soil. *BioMetals* 34, 737–759 (2021).

11. Rehman, M. *et al.* Copper environmental toxicology, recent advances, and future outlook: a review. *Environ. Sci. Pollut. Res.* 26, 18003–18016 (2019).

12. Yruela, I. Transition metals in plant photosynthesis. *Metallomics* 5, 1090–1109 (2013).

13. Schirrmeister, B. E., De Vos, J. M., Antonelli, A. & Bagheri, H. C. Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc. Natl. Acad. Sci.* 110, 1791–1796 (2013).

14. Bernal, M. & Krämer, U. Involvement of Arabidopsis Multi-Copper Oxidase-Encoding LACCASE12 in Root-to-Shoot Iron Partitioning: A Novel Example of Copper-Iron Crosstalk. *Front. Plant Sci.* 12, (2021).

15. Thapa, G., Sadhukhan, A., Panda, S. K. & Sahoo, L. Molecular mechanistic model of plant heavy metal tolerance. *BioMetals* 25, 489–505 (2012).

16. Ghori, N.-H. *et al.* Heavy metal stress and responses in plants. *Int. J. Environ. Sci. Technol.* 16, 1807–1828 (2019).

17. Yruela, I. Copper in plants: acquisition, transport and interactions. *Funct. Plant Biol.* 36, 409–430 (2009).

18. Krzesłowska, M. The cell wall in plant cell response to trace metals: polysaccharide remodeling and its role in defense strategy. *Acta Physiol. Plant.* 33, 35–51 (2011).

19. Ghori, N.-H. *et al.* Heavy metal stress and responses in plants. *Int. J. Environ. Sci. Technol.* 16, 1807–1828 (2019).

20. Kraemer, U. The Dilemma of Controlling Heavy Metal Accumulation in Plants. *New Phytol.* 181, 3–5 (2009).

21. Rai, S., Singh, P. K., Mankotia, S., Swain, J. & Satbhai, S. B. Iron homeostasis in plants and its crosstalk with copper, zinc, and manganese. *Plant Stress* 1, 100008 (2021).

22. Festa, R. A. & Thiele, D. J. Copper: An essential metal in biology. *Curr. Biol.* 21, R877–R883 (2011).

23. Shabbir, Z. *et al.* Copper uptake, essentiality, toxicity, detoxification and risk assessment in soil-plant environment. *Chemosphere* 259, 127436 (2020).

24. Mir, A., Pichtel, J. & Hayat, S. Copper: uptake, toxicity and tolerance in plants and management of Cu-contaminated soil. *BioMetals* 34, 737–759 (2021).

25. Chen, G., Li, J., Han, H., Du, R. & Wang, X. Physiological and Molecular Mechanisms of Plant Responses to Copper Stress. *Int. J. Mol. Sci.* 23, 12950 (2022).

26. Li, X.-Y. *et al.* Physiological and Ultrastructural Responses to Excessive-Copper-Induced Toxicity in Two Differentially Copper Tolerant Citrus Species. *Plants* 12, 351 (2023).

27. Bernal, M. *et al.* Transcriptome Sequencing Identifies SPL7-Regulated Copper Acquisition Genes FRO4/FRO5 and the Copper Dependence of Iron Homeostasis in Arabidopsis[C][W]. *Plant Cell* 24, 738–761 (2012).

28. Kumar, V. *et al.* Copper bioavailability, uptake, toxicity and tolerance in plants: A comprehensive review. *Chemosphere* 262, 127810 (2021).

29. Wairich, A. *et al.* Throwing Copper Around: How Plants Control Uptake, Distribution, and Accumulation of Copper. *Agronomy* 12, 994 (2022).

30. Adrees, M. *et al.* The effect of excess copper on growth and physiology of important food crops: a review. *Environ. Sci. Pollut. Res. Int.* 22, 8148–8162 (2015).

31. Dong, J., Kim, S. T. & Lord, E. M. Plantacyanin Plays a Role in Reproduction in Arabidopsis. *Plant Physiol.* 138, 778–789 (2005).

32. Cai, Y. *et al.* IRON MAN interacts with Cu-DEFICIENCY INDUCED TRANSCRIPTION FACTOR 1 to maintain copper homeostasis. *New Phytol.*

33. Shin, L.-J., Lo, J. & Yeh, K. Copper Chaperone Antioxidant Protein1 Is Essential for Copper Homeostasis1[W][OA]. *Plant Physiol.* 159, 1099–1110 (2012).

34. Leng, X. *et al.* Transporters, chaperones, and P-type ATPases controlling grapevine copper homeostasis. *Funct. Integr. Genomics* 15, 673–684 (2015).

35. Franco, A., Buoso, S., Zanin, L., Pinton, R. & Tomasi, N. Copper Toxicity in Maize: The Severity of the Stress is Reduced Depending on the Applied Fe-Chelating Agent. *J. Plant Growth Regul.* 42, 1567–1581 (2023).

36. Khan, I. *et al.* Identification of novel rice (Oryza sativa) HPP and HIPP genes tolerant to heavy metal toxicity. *Ecotoxicol. Environ. Saf.* 175, 8–18 (2019).

37. Chen, G. & Xiong, S. OsHIPP24 is a Copper Metallochaperone Which Affects Rice Growth. *J. Plant Biol.* 64, 145–153 (2020).

38. Brunetto, G. *et al.* Copper accumulation in vineyard soils: Rhizosphere processes and agronomic practices to limit its toxicity. *Chemosphere* 162, 293–307 (2016).

39. Fagnano, M. *et al.* Copper accumulation in agricultural soils: Risks for the food chain and soil microbial populations. *Sci. Total Environ.* 734, 139434 (2020).

40. Rascio, N. & Navari-Izzo, F. Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Sci.* 180, 169–181 (2011).

41. Pasricha, S. *et al.* Molecular mechanisms underlying heavy metal uptake, translocation and tolerance in hyperaccumulators-an analysis: Heavy metal tolerance in hyperaccumulators. *Environ. Chall.* 4, 100197 (2021).

42. Cobbett, C. Heavy Metals and Plants: Model Systems and Hyperaccumulators. *New Phytol.* 159, 289–293 (2003).

43. Capuana, M. *et al.* Paradoxical effects of density on measurement of copper tolerance in Silene paradoxa L. *Environ. Sci. Pollut. Res.* 25, 1331–1339 (2018).

44. Colzi, I., Pignattelli, S., Giorni, E., Papini, A. & Gonnelli, C. Linking root traits to copper exclusion mechanisms in Silene paradoxa L. (Caryophyllaceae). *Plant Soil* 390, 1–15 (2015).

45. Lin, M. *et al.* The adaptation of root cell wall pectin to copper toxicity in two citrus species differing in copper tolerance: remodeling and responding. *Tree Physiol.* 43, 1825–1840 (2023).

46. Fu, X.-Z. *et al.* Whole-transcriptome RNA sequencing reveals the global molecular responses and ceRNA regulatory network of mRNAs, lncRNAs, miRNAs and circRNAs in response to copper toxicity in Ziyang Xiangcheng (Citrus junos Sieb. Ex Tanaka). *BMC Plant Biol.* 19, 509 (2019).

47. de Oliveira Silva, F. M., de Ávila Silva, L., Araújo, W. L., Zsögön, A. & Nunes-Nesi, A. Exploiting Natural Variation to Discover Candidate Genes Involved in Photosynthesis-Related Traits. in *Photorespiration: Methods and Protocols* (eds. Fernie, A. R., Bauwe, H. & Weber, A. P. M.) 125–135 (Springer, New York, NY, 2017). doi:10.1007/978-1-4939-7225-8_9.

48. Bulut, M., Alseekh, S. & Fernie, A. R. Natural variation of respiration-related traits in plants. *Plant Physiol.* 191, 2120–2132 (2023).

49. Anawar, H. M., Canha, N., Santa-Regina, I. & Freitas, M. C. Adaptation, tolerance, and evolution of plant species in a pyrite mine in response to contamination level and properties of mine tailings: sustainable rehabilitation. *J. Soils Sediments* 13, 730–741 (2013).

50. Kossoff, D. *et al.* Mine tailings dams: Characteristics, failure, environmental impacts, and remediation. *Appl. Geochem.* 51, 229–245 (2014).

51. Thakur, M. *et al.* Metal tolerance in plants: Molecular and physicochemical interface determines the "not so heavy effect" of heavy metals. *Chemosphere* 287, 131957 (2022).

52. MacNair, M. R. & Christie, P. Reproductive isolation as a pleiotropic effect of copper tolerance in *Mimulus guttatus*? *Heredity* 50, 295–302 (1983).

53. Wright, K. M. *et al.* Adaptation to heavy-metal contaminated environments proceeds via selection on pre-existing genetic variation. 029900 Preprint at https://doi.org/10.1101/029900 (2015).

54. Wright, K. M., Lloyd, D., Lowry, D. B., Macnair, M. R. & Willis, J. H. Indirect Evolution of Hybrid Lethality Due to Linkage with Selected Locus in *Mimulus guttatus*. *PLoS Biol.* 11, e1001497 (2013).

55. Jeong, A. Investigating the Genetic, Evolutionary, and Functional Mechanisms of Adaptations to Copper Mine Tailings in *Mimulus guttatus.* (2017).

56. Roberts, S. A. *et al.* Crystal structure and electron transfer kinetics of CueO, a multicopper oxidase required for copper homeostasis in Escherichia coli. *Proc. Natl. Acad. Sci.* 99, 2766–2771 (2002).

57. Gräff, M., Buchholz, P. C. F., Le Roes-Hill, M. & Pleiss, J. Multicopper oxidases: modular structure, sequence space, and evolutionary relationships. *Proteins Struct. Funct. Bioinforma.* 88, 1329–1339 (2020).

58. Djoko, K. Y., Chong, L. X., Wedd, A. G. & Xiao, Z. Reaction Mechanisms of the Multicopper Oxidase CueO from *Escherichia coli* Support Its Functional Role as a Cuprous Oxidase. *J. Am. Chem. Soc.* 132, 2005–2015 (2010).

59. Hoegger, P. J., Kilaru, S., James, T. Y., Thacker, J. R. & Kües, U. Phylogenetic comparison and classification of laccase and related multicopper oxidase protein sequences. *FEBS J.* 273, 2308–2326 (2006).

60. Laubach, Z. M., Murray, E. J., Hoke, K. L., Safran, R. J. & Perng, W. A biologist's guide to model selection and causal inference. *Proc. R. Soc. B Biol. Sci.* 288, 20202815 (2021).

61. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48 (2015).

62. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* 4, 1686 (2019).

63. Selby, J. P. & Willis, J. H. Major QTL controls adaptation to serpentine soils in *Mimulus guttatus*. *Mol. Ecol.* 27, 5073–5087 (2018).