

IDENTIFICATION OF CORE GENOMIC MECHANISMS OF ABIOTIC STRESS  
RESPONSE IN GRASSES

By

Anna Pardo

A DISSERTATION

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

Horticulture – Doctor of Philosophy

2024

## ABSTRACT

Abiotic stress is any deviation from optimum growth conditions for plants caused by non-biological factors such as water, temperature, nutrients, salts, or light. Such stresses have been important drivers of plant adaptation in ecological settings and are important constraints on global agricultural yield. Climate change is projected to increase the frequency and severity of abiotic stress events in the future, including extreme weather events such as droughts and temperature fluctuations, as well as desertification and soil salinization, which will negatively impact agriculture. Many major crops, such as maize, wheat, rice, barley, and sorghum among others, are found in the grass family, Poaceae, which contains over 11,000 species and is also highly ecologically important. In particular, the subfamily Chloridoideae is highly resilient and contains most of the desiccation tolerant grasses. Understanding the mechanisms of abiotic stress response in both cereal crops and naturally resilient grasses is therefore important for improvement of agricultural resilience in the face of climate change.

Each chapter of this dissertation examines abiotic stress response in one or more grass species from a genomics perspective. Chapter 1 gives an overview of abiotic stress responses and adaptations in the grasses, including desiccation tolerance, the ability to survive near-complete drying, which is found in 40 grass species. Chapter 1 also discusses the status of genome sequencing in the grasses; approximately 1% of the species in the family have at least one reference genome, with several important crops such as maize and wheat having multiple reference genomes for different genotypes. Chapter 2 is a meta-analysis of approximately 1,900 RNA-sequencing samples for six different stress conditions in maize: drought, salt, cold, heat, flooding, and low nitrogen. The goal of this study was to find core stress-responsive genes via two methods: set operations and random forest classification. We found that these two methods identified largely distinct sets of core genes; random forest identified core genes that were important for predicting if a sample was stressed or control, sometimes regardless of whether those genes were differentially expressed in any stress condition. Furthermore, core genes were enriched in transcription factors generally as well as specific families, such as bZIP, NAC, HSF, and ERF. We hypothesized that these transcription factors may regulate other core genes as well as stress-specific genes.

In Chapter 3, the genome of desiccation tolerant grass *Eragrostis nindensis* was improved by high-fidelity long read sequencing. The new assembly is 20-fold more contiguous than the

previous assembly, with a contig N50 of over 10 Mb for 828 contigs. The *E. nindensis* genome, already known to be tetraploid, was found to be likely autotetraploid using this improved assembly and annotation. In Chapter 4, the improved *E. nindensis* genome along with 85 others was used to find groups of orthologous genes across the grass phylogeny. Subsequently, expanded orthogroups were found for desiccation tolerant compared to sensitive Chloridoideae species, as well as conserved transcription factor-binding motifs in desiccation tolerant chloridoids. We thus concluded that both expansion of certain gene families (early light-induced proteins, thaumatin-like proteins, expansin precursors, and others) and gene expression regulatory changes (recruitment of motifs for TCP, BBR/BPC, and TIFY family transcription factors) have contributed to the evolution of desiccation tolerance in the Chloridoideae.

## ACKNOWLEDGEMENTS

As has been written in many dissertations before mine, the completion of a PhD is not a solo endeavor. There are many people I would like to thank for their help and support along the way. Firstly, my advisor, Dr. Bob VanBuren, whose guidance has been invaluable throughout this journey. He is truly the best advisor I could have asked for: not only a brilliant scientist, but a caring person as well, who always emphasized the importance of balance and never questioned my occasional (well, perhaps more than occasional) need for a mental health day. Bob has always helped me shape my semi-coherent ideas into solid science, and his flexibility was invaluable when I decided to complete the last year and a half of my PhD work remotely from a different state. As a first-generation grad student, it's also been amazing to have a first-generation advisor. Bob, I can't tell you what an impact you've made on my life as well as my career. Thank you so much.

I also have some wonderful committee members. Many thanks to Dr. Jianrong Wang of the CMSE department for his invaluable technical feedback when I waded into the world of machine learning. Dr. Addie Thompson of PSM was immensely helpful when I started working on maize, doing everything from providing seeds for an ill-fated drought priming experiment to staying after a committee meeting with me to discuss analysis options for my core stress project. Last but not least, Dr. Pat Edger of HRT has generously contributed his comparative genomics expertise. All my committee members have always been supportive and concerned for me as a person, which I greatly appreciate. Not every PhD student is so lucky and I am thankful to have had such a great committee.

I have been fortunate to work with some true quality people and absolute superstar scientists throughout my time in the VanBuren lab. Dr. Jennifer Wai helped me find my way around the lab when I first started and has been a good friend full of helpful scientific suggestions ever since. Dr. Brian St. Aubin laid the groundwork on chromatin openness studies in *Oropetium* and *E. nindensis*. Dr. Kevin Bird and Dr. Jeremy Pardo (more on him later) were the first students to graduate from the VanBuren lab; Jeremy in particular has advised me on science every step of my PhD. Dr. (soon to be Professor!) Rose Marks has been a wonderful friend and mentor, and very generous with her incredible desiccation tolerance knowledge. Dr. Ian Gilman has provided both technical knowledge and emotional support over the last year, and I look forward to collaborating with him in future.

My fellow June 2024 Defenders, McKena Wilson and Serena Lotreck, have been wonderfully supportive throughout my dissertation writing process, as well as being amazing friends for the last five years. The last semester of a PhD is a heck of a time, but having you two amazing scientists and science buddies there doing it with me has been so much better than doing it alone would've been. I really wish I could see you defend in person, but I'll be cheering for you over Zoom, and wherever you end up in your careers I'll be cheering for you too. Just remember, if you have anxiety: throw it out the window!

I'm also grateful to younger grad students of the VanBuren lab, my friends and sometimes informal mentees. Jenny Schuster is an incredible scientist with whom I've bonded over the pain of tissue culture and the beauty of Maine, among other things. I'm so excited to see what you do in the future, Jenny, whether at NASA or elsewhere! Along with Jenny, Cathy Mercado has brought molecular biology back to the VanBuren lab, and I'm so looking forward to the results of her long-term desiccation experiment. I'm really grateful I've gotten to hear about her work in DT right from the beginning. Although I've known Maddy Creach and Elliot Braun for less time, I've really enjoyed having them in the lab and look forward to reading their work as well. I also want to thank the two undergraduate students, Kirk Maibach and Cate Kirkwood, who helped out with my ill-fated drought priming experiment in spring 2022. I wouldn't have figured out that I had no phenotype nearly as quickly without their help.

I would be remiss if I didn't acknowledge all the other grad school friends I've had outside of my lab. The whole of HOGS and IMPACTS are great student groups, but in particular I want to call out Dr. Charity Goeckeritz, Dr. Kathleen Rhoades, their husbands Indy Uribe and Sam Rhoades (not students but still friends), and my first-year friends Scott Teresi and former roommate Christina Chiu. In addition to their friendship (and invaluable moving help!), Charity and Kathleen gave great technical assistance when I was annotating the *E. nindensis* genome and looking into its polyploid origin. In particular, without Charity, I never would have been able to run MAKER or AUGUSTUS, so many thanks to her!

Speaking of technical assistance, many thanks to Dr. Kevin Childs of the MSU genomics core for his help with Nanopore RNA-seq data processing. Thanks also to Jim Klug and Cody Keilen of the MSU growth chamber facility and Dr. Chrislyn Particka, director of the research greenhouses. These growth facilities were invaluable in maintaining the plants for my dissertation, and I'm grateful to their directors for keeping them running. For funding, I have to

thank the MSU graduate school for their University Enrichment Fellowship, the NRT-IMPACTS program, and the NSF. Endless thanks to Jyothi Kumar, the administrator of IMPACTS, and the hort office staff, especially Sherry Mulvaney and Meghan Hill (what would we do without you?), for their answers to my random questions throughout my program.

I would not be here at the end of my PhD program without all the people who supported me through my undergraduate work at the University of New Hampshire. First off, my research mentor, Dr. Subhash Minocha, academic advisor, Dr. Estelle Hrabak, and favorite teaching professor, Dr. Tom Davis, advised me throughout my program, taught some of my favorite classes, and wrote way too many letters of recommendation for me without a word of complaint. I am endlessly grateful to them, especially to Dr. Minocha, who at times believed in me more than I believed in myself. My friends from the Minocha lab, especially Sefali and Chandra, helped build my research skills and always treated me as an equal even though I was just an overachieving undergrad. Very special thanks to the UNH McNair Scholars Program staff of 2017-2019, Selina Choate and Tammy Gewehr, for research and travel funding, emotional support, and mentorship through graduate school applications. Thanks also to the McNair summer 2017 cohort and peer mentors for being such a great friend group of fellow low-income, first-generation, and/or underrepresented scholars. I also have to thank the Hamel Center for Undergraduate Research and its director, Dr. Paul Tsang, for awarding me summer and semester research funding as well as multiple travel grants. Many thanks also to the UNH greenhouse staff, especially Luke Hydock, for helping me figure out how to grow rice and putting up with my weird watering schedules every time I did a stress experiment for two and a half years. My undergrad was a great experience thanks to all these people and more.

I've been fortunate to have two great friends from undergrad who I still keep in contact with, Bethynie Cooper and (future Dr.) Emily Berry. From riding elevators in fancy hotels in Atlanta, to getting lost on San Francisco public transit, we did some great bonding during our McNair summer. I'm so glad we've kept in touch even though we're all graduated from UNH now. I have loved catching up at our monthly Zoom meetings, and I'm now really enjoying our D&D game - don't lick the purple mist! You two have been so supportive throughout my PhD and I can't wait for our next adventures together, whether virtual or in person.

My oldest friend, Olivia Hancock, has been there for me for the last 20-plus years, ever since we met at a homeschool get-together when I was six and she was eight. Even though we

took very different paths in life, she has always been there to support me and listen to my venting about life as a scientist. I am so, so grateful for her steadfast sisterhood all these years. Thank you, Olivia. I'm so lucky to have you as a friend.

I'm fortunate to have some amazing in-laws in the Pardo family. While I will be a first-generation PhD graduate, there are many PhD Pardos who have gone before me and let me know that what I'm doing is possible. Dr. Mickey Pardo and Dr. Yudi Pardo, my brothers-in-law, and Dr. Scott Pardo, my father-in-law, are all scientists themselves and have understood what I'm going through as I finish up my PhD. Scott in particular has always been there to give me statistical advice, which has been immensely helpful in developing my analyses. My mother-in-law, Gail Pardo, has not done a PhD herself but really understands what it's like and has always been supportive of me and encouraged me when I don't believe that I can finish. I'm so grateful that all of the Pardos have welcomed me into their *famiya*; I feel very blessed to have them in my life.

I cannot thank my own family enough for all they have done for me throughout my life. My parents, Bruce and Jill Haber, have really done everything they can for me, from homeschooling me K-12 so I could tailor my education to my own interests, to continuing to house me so I could save a lot of money by commuting to my undergrad, to helping me move to Michigan and then back to Massachusetts. All throughout, they have been supportive of my scientific dreams and increasing educational goals, even though neither of them graduated from college and they didn't always understand exactly what I did for work. They've always been excited to hear about my adventures and spend time with me whenever possible, and always encouraged me to take good care of my mental health. I really can't thank them enough.

My extended family, in particular both sets of my grandparents, Stan and Patricia Haber and David and Eleanor Bellemore, as well as my Aunt Theresa and Uncle David Gilfoy and Uncle Dennis and Aunt Tricia Bellemore, have always been supportive as well. I'm so grateful for their love and kindness throughout my life. Not going to lie, a major motivation for me to keep going on days when I wanted to quit this PhD was the thought of how proud they'd be of having a doctor in the family for the first time! Family, I know I'm technically a Pardo now, but feel free to call me Dr. Haber whenever you want.

My brother Peter Haber, MS, is, I suppose, my real oldest friend, since I've known him ever since he was born. Once I got over youthful sibling rivalry (sorry, Peter), I really enjoyed

having him as a homeschool classmate, playmate, friend, and later, fellow scientist. It's been great visiting with him both in Ohio and Maine, and I'm so proud of him for getting an advanced degree before I did - as I like to joke, the only thing my younger brother ever did first in his life! I can't thank him enough for his love, support, friendship, and the fun we've had throughout our lives.

My husband, Dr. Jeremy Pardo, has been my greatest support throughout my PhD. I'm so thankful that we didn't remain just labmates over the last four and a half years. He is my best collaborator as well as my life partner, and without his technical assistance (so extensive he became a coauthor on every one of my research chapters), this dissertation absolutely would not be as high quality as it is today. He has supported me in every way so I could make it to this finish line and beyond, from making me dinner to giving hugs whenever needed, to actually helping run analyses. Thank you so much, Jeremy. I love you forever.

Finally, I cannot express my gratitude to the Creator of all that I study, the Giver of every blessing in my life. May I glorify You always.



## TABLE OF CONTENTS

<b>Chapter 1: Introduction</b> .....	1
REFERENCES.....	10
APPENDIX.....	16
<b>Chapter 2: Stress-responsive transcription factor families are key components of the core stress response in maize</b> .....	20
REFERENCES.....	38
APPENDIX A: FIGURES AND TABLES.....	43
APPENDIX B: SUPPLEMENTAL FIGURES AND TABLES.....	53
<b>Chapter 3: Improved genome assembly and annotation for <i>Eragrostis nindensis</i>, an autotetraploid desiccation tolerant grass</b> .....	64
REFERENCES.....	71
APPENDIX.....	75
<b>Chapter 4: Genomic signatures of desiccation tolerance in the resilient grass subfamily Chloridoideae</b> .....	79
REFERENCES.....	92
APPENDIX A: FIGURES AND TABLES.....	103
APPENDIX B: SUPPLEMENTAL FIGURES AND TABLES.....	111
<b>Future Directions</b> .....	122
REFERENCES.....	125

## Chapter 1: Introduction

### *Agricultural and ecological impacts of abiotic stress*

Abiotic stresses such as drought and desiccation, salt, temperature extremes including freezing, flooding, nutrient deficiency, excess heavy metals, and low and high light intensities can all have negative impacts on plants both in agricultural and natural settings. Abiotic factors have long been a major driver of plant adaptation and evolution; for example, desiccation tolerance was essential for the colonization of land by charophytic algae (Oliver et al., 2000). Furthermore, abiotic stress is a major constraint on global agricultural yield. In different regions of the world, heat, cold, salt, flooding, and drought events are projected to increase due to climate change (Oshunsanya et al., 2019). This means it is important to understand how plants respond to and tolerate abiotic stresses in order to make more resilient crops for a changing climate.

### *Grasses and their stress adaptations*

The grass family, Poaceae, is one of the most agriculturally and ecologically important plant families in the world. It contains 11,783 species including the top three global crops: rice (*Oryza sativa*), wheat (*Triticum aestivum*), and maize (*Zea mays*). Grasses originated in tropical areas and radiated to cover approximately 40% of the world's land surface across virtually every biome (Strömberg, 2011). The grasses are divided into two major clades named for the subfamilies they contain, the BOP (Bambusoideae, Oryzoideae, and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae), although there are three additional early-diverging subfamilies that contain only a few dozen species (Huang et al., 2022; Soreng et al., 2022). Most BOP clade grasses are cool-season species that utilize C3 photosynthesis; most if not all C4 species in the family are in the PACMAD clade (Pardo and VanBuren, 2021). These clades show clear differences in distribution as well as stress adaptation. For instance, the BOP clade is generally distributed in temperate and subtropical regions, and Pooideae species dominate environments with cold winters (Schubert et al., 2020). PACMAD grasses are generally tropical and highly adapted to heat and drought, although certain PACMAD species such as *Phragmites australis* (Arundinoideae) and *Danthonia decumbens* (Danthonioideae) have adapted to cooler environments (Schubert et al., 2020; Pardo and VanBuren, 2021).

Poaceae contains species utilizing both C3 and C4 photosynthesis. In *Alloteropsis semialata* (Panicoideae), which is the only known species containing both C3 and C4 genotypes, the use of C4 photosynthesis was found to broaden the plant's ecological niche (Lundgren et al., 2015). However, because the C4 pathway mostly evolved in the tropical PACMAD grasses, it is typically associated with tolerance to conditions such as drought and heat, although C4 photosynthesis alone is not sufficient for drought tolerance as indicated by the drought sensitivity of C4 crops like maize (Pardo and VanBuren, 2021). Drought-tolerant C4 panicoid and chloridoid grasses have other adaptations leading to their high water use efficiency. These include dumbbell-shaped stomatal guard cells that can respond more quickly to changes in water potential, as well as stomatal distribution either on the abaxial side or on both sides of the leaf that reduces evapotranspiration (Pardo and VanBuren, 2021). In addition, some grasses have unique adaptations to cold stress. Under cold as well as other stresses, it is common for grass cells to accumulate sugars as osmoprotectants (see below). In cold-adapted taxa such as the Pooideae, fructans are accumulated in place of other sugars; this is a key cold adaptation which reduces feedback inhibition of photosynthesis by high levels of sucrose and other sugars (Schubert et al., 2020).

While many grasses are broadly tolerant to stresses such as drought, heat, and cold, some are highly tolerant to other stressors, such as desiccation and salt. Desiccation tolerance (DT) is the ability to survive near-complete water loss, essentially an extreme form of drought tolerance. While common in angiosperm seeds and necessary for the initial movement of plants onto land, DT is uncommon in angiosperm vegetative tissues (Oliver et al., 2000). Vegetative DT has, however, convergently re-evolved in various angiosperm lineages, most likely through rewiring of seed DT pathways (Oliver et al., 2000). DT has evolved repeatedly within the grass family, and there are forty DT grasses, mostly in Chloridoideae with some representation in Micrairoideae and two species in Pooideae (Table A1.2) (Marks et al., 2021a; Pardo and VanBuren, 2021). Within the Chloridoideae, there are 17 DT species that have been identified in tribe Cynodonteae, 7 in the Eragrostideae, and 7 in the Zoysieae (Table A1.2). Of the DT grasses, *Sporobolus stapfianus* (Zoysieae), *Eragrostis nindensis* (Eragrostideae), *Tripogon loliiformis* (Cynodonteae), and *Oropetium thomaeum* (Cynodonteae) are among the fifteen best-studied DT angiosperm species (Tebele et al., 2021). *E. nindensis* is the only well-studied DT species in the same genus as a cereal crop, teff (*E. tef*), which is a staple in Ethiopia and Eritrea

(Pardo et al., 2020), but given the importance of grasses in agriculture generally, lessons learned from all DT grass species may be relevant for future crop improvement.

Similar to DT, salt tolerance has evolved multiple times in the grasses, including multiple independent evolutions within Chloridoideae (Bennett et al., 2013). This includes the halophyte grasses, i.e. those that can complete their life cycle in environments with high salt concentrations, as well as salt-tolerant glycophytes, i.e. species that do not complete their life cycle in high salt environments but have a transient tolerance to salt (Roy and Chakraborty, 2014). In most cases, the latter use salt glands on the leaves to remove excess ion buildup in the tissues, while halophytes simply exclude ions from their tissues altogether (Roy and Chakraborty, 2014). Salt glands and salt tolerance are found in many species in the PACMAD clade compared to other grass lineages (Pardo and VanBuren, 2021). About 8% of the world's halophyte angiosperms are grasses (Roy and Chakraborty, 2014).

#### *Availability of genomic resources for the grasses*

The order Poales, to which Poaceae belongs, is one of the most sequenced orders in the plant kingdom (Marks et al., 2021b). There are 114 total grass species (0.97% of the family) whose genomes have been sequenced, including species such as rice, wheat, and maize which have multiple genotypes with reference-quality genomes. Eight of the twelve subfamilies have at least one sequenced species (Table A1.1). Panicoideae and Pooideae are the most highly sequenced subfamilies with 33 sequenced species each, followed by Oryzoideae with 21 sequenced species, all in the tribe Oryzeae, which is the most highly sequenced tribe in the Poaceae (Table A1.1). It is notable that these three subfamilies contain the top three global crops, maize, wheat, and rice, respectively. Other subfamilies have fewer sequenced species: in descending order, Chloridoideae with 17 sequenced species, Bambusoideae with 7, and Pharoideae, Anomochlooideae, and Arundinoideae with 1 each (Table A1.1). Within the Chloridoideae, members of all three of the largest tribes, Cynodonteae, Eragrostideae, and Zoysieae, have been sequenced (Table A1.1).

In addition to taxonomic coverage, it is important to sequence stress-resilient species so there are resources available for the study of their tolerance mechanisms. Sequenced grass species with vegetative DT include *Oropetium thomaeum*, *O. capense*, *Eragrostis nindensis*, *Sporobolus stapfianus*, *Tripogon minimus*, and *Microchloa caffra*, six in total, or 15% of DT grass species (Table A1.3) (VanBuren et al., 2018; Pardo et al., 2020; Chávez Montes et al.,

2022; Marks et al., 2024). There are also genomes for desiccation sensitive (DS) sister taxa in two of these genera, namely *E. tef*, *E. curvula*, and *S. pyramidalis* (Carballo et al., 2019; VanBuren et al., 2020; Chávez Montes et al., 2022), which have previously been used in experimental comparisons with their DT counterparts. There are, however, many other DT grass species without published genomes, including *T. loliformis*, one of the most studied DT angiosperms (Tebele et al., 2021).

Of the 114 sequenced grass species, 25 have been identified as salt tolerant by (Bennett et al., 2013) (see Table A1.3 for species list and genome references). Additionally, *Panicum virgatum* (switchgrass) has been sequenced, although the genotype sequenced was AP13 (Lovell et al., 2021), rather than salt tolerant var. *cubense* (Bennett et al., 2013); thus, it is not included in Table A1.3. Since there are 200 total salt-tolerant grass species identified by (Bennett et al., 2013), it follows that 12.5% of salt tolerant grasses have sequenced genomes, a respectable number. However, given the general under-representation of wild plants among species with sequenced genomes (Marks et al., 2021b), it would be beneficial to sequence still more salt-tolerant grass species. The same applies to DT species as discussed above.

In addition to sequenced genomes, there are many gene expression datasets available using both microarray and RNA-sequencing technology for various grass species. On NCBI, there are 39,938 “transcriptome or gene expression” BioProjects for all land plants. A search for Poaceae reduces this to 16,194 BioProjects, meaning that about 40.5% of all plant transcriptome BioProjects on NCBI are for grasses. Rice has 2,196 transcriptome BioProjects, maize 2,137, and wheat 464. The top three cereals account for 29.6% of all grass transcriptome BioProjects. This high availability of transcriptomic data makes grasses a useful family for studying various plant traits, including stress resilience. In addition, meta-analysis of these public data is an attractive avenue of research, especially for particularly well-resourced species such as rice and maize.

#### *Cellular responses to abiotic stresses*

Abiotic stress leads to multiple responses on the cellular and molecular level within plants. One of the most important of these is inhibition of photosynthesis, which occurs under most abiotic stresses including drought, salt, heat, and cold among others (Singh and Thakur, 2018; Muhammad et al., 2021). Although the mechanism of negative impact on photosynthesis, such as carbon dioxide limitation due to stomatal closure, may be stress-specific, the ultimate result is typically that the rate of the light-independent reactions (Calvin-Benson-Bassham cycle)

is reduced, while the light-dependent reactions are comparatively unaffected. This leads to an excess of excited chlorophyll and the production of reactive oxygen species (ROS), which include molecules such as superoxide, the hydroxyl radical, and hydrogen peroxide among others (Choudhury et al., 2017). Excess ROS from photosynthetic inhibition can be quenched by other pigments, including carotenoids (Singh and Thakur, 2018).

ROS serve dual functions under stress: a beneficial signaling function and a detrimental oxidative stress function. Catabolism of the higher polyamines, spermidine and spermine, has been linked to the generation of signaling ROS (Minocha et al., 2014). ROS production in the chloroplast can lead to programmed cell death (Choudhury et al., 2017). Additionally, excess ROS are generally detrimental as they can oxidize membranes, macromolecules such as proteins, and other cellular components. Therefore, an important aspect of stress response in plants is antioxidant activity. ROS can be scavenged by molecules such as polyamines, glutathione, and ascorbate, in addition to pigments, and enzymes such as superoxide dismutase, ascorbate peroxidase, and glutathione reductase are also important for this process. The gene expression and catalytic activity of these enzymes often increases under stress (Choudhury et al., 2017).

Many plant abiotic stress responses are aimed at protection of cellular components and macromolecules, including membranes, proteins, and DNA, from damage. For example, under stresses with an osmotic component such as water deprivation (both drought and desiccation), salt, and cold (both chilling and freezing), many plant species accumulate small molecules known as osmoprotectants or compatible solutes, including sugars, amino acids, betaines, and polyamines (Kumar et al., 2018). These molecules act to encourage cellular water retention in unfavorable conditions without being toxic to the cell (Verslues et al., 2006), and sometimes as “water replacers” to stabilize macromolecules such as proteins, especially in the case of sugars under desiccation (Hoekstra et al., 2001).

Polyamines are small, cationic, aliphatic molecules containing multiple amino groups. They are essential for life in all kingdoms, including plants. The main polyamines in plants are putrescine (a diamine), spermidine (a triamine), and spermine (a tetraamine), although other polyamines also occur with less frequency. Polyamines have long been linked to abiotic stress response in plants, both in beneficial and detrimental ways depending on the species and particular polyamine in question (Minocha et al., 2014; Tyagi et al., 2023). Polyamines have frequently been found to be stress-protective, as osmoprotectants, signaling components, and

binders of macromolecules such as DNA; many studies have found that exogenous application of polyamines or overexpression of their biosynthetic genes leads to higher stress tolerance (Minocha et al., 2014; Tyagi et al., 2023). However, their catabolism can also lead to ROS such as hydrogen peroxide, which in excess can have negative impacts on stressed cells (Minocha et al., 2014). Furthermore, in certain species, some polyamines have been linked to damage from cold stress (Tyagi et al., 2023). Thus, polyamines can be beneficial or detrimental under stress, depending upon the species and the polyamine.

Various sugars can act as compatible solutes/osmoprotectants under abiotic stress, including sucrose, fructose, glucose, fructans, and raffinose family oligosaccharides (RFOs) (Kumar et al., 2018). In DT plants, the importance of sugars as osmoprotectants and water replacers has long been recognized (Zhang et al., 2016). During desiccation, sugars such as trehalose and others aid in the vitrification of cells as water is lost, a key response in DT plant species (Zhang et al., 2016; Kumar et al., 2018).

Many of the abiotic stress responses discussed here are regulated by the activity of various phytohormones. Some of the most important of these are abscisic acid (ABA), jasmonic acid (JA)/jasmonates, ethylene, and salicylic acid (SA). ABA is widely recognized as one of the most important hormones for abiotic stress response. ABA signaling occurs through the MAP kinase (MAPK) pathway and leads to stomatal closure under osmotic stress, which reduces water loss (de Zelicourt et al., 2016; Dar et al., 2017). JA can also regulate stomatal closure under drought and herbivore stress, as well as mediating other responses such as ROS scavenging (Wang et al., 2020b). Ethylene, which has been particularly noted as important under flooding but is also active in other stress conditions, acts to inhibit growth and improve stress tolerance through changing the expression of ERF transcription factors (Chen et al., 2022). Under various abiotic stress conditions in different plant species, SA has been found to increase antioxidant activity (Zaid et al., 2021). These and other phytohormones do not act independently, but rather there is significant crosstalk among hormone signaling pathways. In addition, exogenous application of phytohormones has been shown to improve tolerance to multiple different stresses in multiple species of plants.

#### *Stress-responsive transcription factors and abiotic stress expression regulation*

The cellular abiotic stress responses discussed above are typically mediated by changes in gene expression, which is subject to regulation by various families of transcription factors (TFs).

Depending on the particular TF, regulation may take the form of activation or repression, usually via binding to a motif in a promoter or enhancer region upstream of the gene being regulated. There are a large number of different stress-responsive TF families which are involved in both hormone-dependent and independent signaling pathways. These include the MYB, bZIP, MYC, NAC, AP2/ERF, WRKY, bHLH, and BBR/BPC families among others, all of which have been shown to be stress-responsive.

The APETALA2/ETHYLENE RESPONSE FACTOR (AP2/ERF) TF family is large and diverse, with several subfamilies, the most important of which are AP2, ERF, and C-REPEAT BINDING FACTOR/DEHYDRATION-RESPONSIVE ELEMENT BINDING (CBF/DREB) (Yoon et al., 2020; Ma et al., 2024). The CBF/DREB genes are mainly responsive to and involved in water deficit and low temperature stresses (Jan et al., 2017), but can also be involved in growth regulation via regulation of gibberellic acid synthesis (Ma et al., 2024). The ERFs are responsive to ethylene, which is particularly important under flooding conditions (Tewari and Mishra, 2018; Ma et al., 2024), but also under other stress conditions as discussed above.

The MYB TF family has four subfamilies, R1-MYB, R2R3-MYB, R1R2R3-MYB, and 4R-MYB, depending on which repeats they have (Yoon et al., 2020). Various MYBs are activators in the ABA signaling pathway under drought and salt stress as well as during seed germination (Wang et al., 2021b). Certain MYBs, for example MYB15 in *Arabidopsis*, also work to activate CBF/DREB expression under cold stress (Wang et al., 2021b). In the process of gene regulation, MYBs also interact with other TFs of different families, such as MYCs and WRKYs (Yoon et al., 2020). Along with the WRKY, C2C2, bZIP, and bHLH families, the MYB family was noted as one of the families with the most desiccation- and rehydration-responsive members in DT eudicot *Myrothamnus flabellifolia* (Ma et al., 2015). Furthermore, an R2R3 MYB has been identified as a key negative regulator involved in DT in the moss *Syntrichia ruralis* (Zhang et al., 2024). Therefore, MYBs are important regulators of response to multiple stresses.

Other repressor TFs involved in abiotic stress include the BARLEY B-RECOMBINANT/BASIC PENTACYSTEINE (BBR/BPC) family, which recruits the polycomb repressor complex to negatively regulate its targets (Sahu et al., 2023). The targets of the BBR/BPCs include developmental genes such as the TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) TFs, which in turn



regulate other developmental genes (Li, 2015), as well as gibberellic acid oxidases, among others (Sahu et al., 2023). They are also active in abiotic stress, and their repression can work both for and against stress tolerance, depending on the particular family member. For example, BPC1 and BPC2 increase salt tolerance via repression of GALACTAN SYNTHASE 1 (Yan et al., 2021a; Sahu et al., 2023), but BPC2 also acts as a repressor of a late embryogenesis abundant (LEA) protein, which is a type of macromolecular chaperone involved in drought and cold tolerance (Sahu et al., 2023). BPC6 is involved in regulation of cuticular wax biosynthesis (Sahu et al., 2023). The BBR/BPC family is an important example of the importance of both activation and repression as regulatory modes for abiotic stress tolerance.

The BBR/BPCs are not the only repressor TFs important for abiotic stress response. bHLH122 acts as a repressor of an ABA catabolism gene (Khan et al., 2018), which can be detrimental under stress conditions; thus, bHLH122 acts to improve stress tolerance. There are various other TF families that interact with ABA signaling. For instance, a subfamily of the bZIPs, the ABA-responsive element binding factors (ABFs), is activated by ABA signal transduction and goes on to activate members of the AP2/ERF and NAM/ATAF/CUC2 (NAC) families, although only some NACs are ABA-responsive (Marques et al., 2017; Yoon et al., 2020). Activated NACs, in turn, regulate ABA biosynthesis, specifically the *NCED3* gene, as well as other stress-related genes including LEAs, COLD-RESPONSIVE (COR) genes, and CBF/DREBs, depending on the particular NAC family member involved (Marques et al., 2017). Thus, these TFs' activities downstream of initial ABA signaling leads to stress response and tolerance in plants.

ABA is, of course, not the only important hormone under abiotic stress; as discussed above, JA is important as well. MYC2, another basic helix-loop-helix (bHLH) gene, has been noted for its interaction with JA signaling. It is normally suppressed by JAZ proteins, but when JA is present, these are degraded, enabling MYC2 to activate antioxidant genes. This protein is also activated by ABA (Yoon et al., 2020). JA also activates members of the bHLH family (Yoon et al., 2020). The WRKY family, which is often involved in developmental regulation during stress (Yoon et al., 2020), is also involved in ABA-related pathways including stomatal closure during drought (Li et al., 2020). WRKYs have also been linked to desiccation tolerance in DT eudicots (Wang et al., 2009; Ma et al., 2015).

Heat shock factors (HSFs) are the direct regulators of heat shock proteins (HSPs), a class of molecular chaperones (Andrási et al., 2021). HSFs have mostly been implicated in heat response, but also act under other stress conditions, including drought and salt (Guo et al., 2016). There are two subfamilies of HSFs, HSFAs and HSFBs; HSFAs can directly bind DNA and regulate their targets' transcription, and their overexpression has been shown to improve stress tolerance in multiple cases, while HSFBs lack DNA-binding domains and are thought to be coactivators of HSFAs (Guo et al., 2016). Along with the other TF families discussed here, HSFs are important regulators of stress-important genes.

### *Conclusion*

Abiotic stress response in grasses is a multifaceted phenomenon. It can be studied through the lens of evolution and tolerance adaptations, such as desiccation tolerance in the Chloridoideae, or at the cellular and gene expression levels, such as commonalities in response pathways among different abiotic stresses. At the organismal level, grasses are a good system for the study of abiotic stress due to their various resilience adaptations and excellent genomic resources. At the cellular and molecular level, reactive oxygen species and osmoprotectants are major players which are regulated by phytohormone signaling pathways and many important transcription factor families, both activators and repressors. In this dissertation, we study the regulation of abiotic stress response in maize and desiccation tolerant chloridoid grasses by generating new genomic resources as well as leveraging those that are publicly available.

## REFERENCES

- Andrási N, Pettkó-Szandtner A, Szabados L** (2021) Diversity of plant heat shock factors: regulation, interactions, and functions. *J Exp Bot* **72**: 1558–1575
- Beier S, Himmelbach A, Colmsee C, Zhang X-Q, Barrero RA, Zhang Q, Li L, Bayer M, Bolser D, Taudien S, et al** (2017) Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci Data* **4**: 170044
- Bennett TH, Flowers TJ, Bromham L** (2013) Repeated evolution of salt-tolerance in grasses. *Biol Lett* **9**: 20130029
- Cai L, Comont D, MacGregor D, Lowe C, Beffa R, Neve P, Sasaki C** (2023) The blackgrass genome reveals patterns of non-parallel evolution of polygenic herbicide resistance. *New Phytol* **237**: 1891–1907
- Carballo J, Santos B a. CM, Zappacosta D, Garbus I, Selva JP, Gallo CA, Díaz A, Albertini E, Caccamo M, Echenique V** (2019) A high-quality genome of *Eragrostis curvula* grass provides insights into Poaceae evolution and supports new strategies to enhance forage quality. *Sci Rep* **9**: 10250
- Chakravartty N, Randowski L, Pirro S** (2023) The Complete Genome Sequence of *Cymbopogon citratus* (Poaceae, Poales), Lemon Grass. *Biodivers Genomes* **2023**: 28–29
- Chávez Montes RA, Haber A, Pardo J, Powell RF, Divisetty UK, Silva AT, Hernández-Hernández T, Silveira V, Tang H, Lyons E, et al** (2022) A comparative genomics examination of desiccation tolerance and sensitivity in two sister grass species. *Proc Natl Acad Sci* **119**: e2118886119
- Chen H, Bullock DA, Alonso JM, Stepanova AN** (2022) To Fight or to Grow: The Balancing Role of Ethylene in Plant Abiotic Stress Responses. *Plants* **11**: 33
- Chen S, Du T, Huang Z, He K, Yang M, Gao S, Yu T, Zhang H, Li X, Chen S, et al** (2024) The *Spartina alterniflora* genome sequence provides insights into the salt-tolerance mechanisms of exo-recretohalophytes. *Plant Biotechnol J*. doi: 10.1111/pbi.14368
- Choudhury FK, Rivero RM, Blumwald E, Mittler R** (2017) Reactive oxygen species, abiotic stress and stress combination. *Plant J* **90**: 856–867
- Dar NA, Amin I, Wani W, Wani SA, Shikari AB, Wani SH, Masoodi KZ** (2017) Abscisic acid: A key regulator of abiotic stress tolerance in plants. *Plant Gene* **11**: 106–111
- Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, Yang X, Peng Q, Wang Y, Jia L, et al** (2017) *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat Commun* **8**: 1031

- Guo M, Liu J-H, Ma X, Luo D-X, Gong Z-H, Lu M-H** (2016) The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. *Front Plant Sci.* doi: 10.3389/fpls.2016.00114
- Hoekstra FA, Golovina EA, Buitink J** (2001) Mechanisms of plant desiccation tolerance. *Trends Plant Sci* **6**: 431–438
- Huang W, Zhang L, Columbus JT, Hu Y, Zhao Y, Tang L, Guo Z, Chen W, McKain M, Bartlett M, et al** (2022) A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C4 photosynthesis. *Mol Plant* **15**: 755–777
- Jan AU, Hadi F, Ahmad A, Rahman K** (2017) Role of CBF/DREB Gene Expression in Abiotic Stress Tolerance. A Review.
- Kang S-H, Kim B, Choi B-S, Lee HO, Kim N-H, Lee SJ, Kim HS, Shin MJ, Kim H-W, Nam K, et al** (2020) Genome Assembly and Annotation of Soft-Shelled Adlay (*Coix lacryma-jobi* Variety *ma-yuen*), a Cereal and Medicinal Crop in the Poaceae Family. *Front Plant Sci.* doi: 10.3389/fpls.2020.00630
- Khan S-A, Li M-Z, Wang S-M, Yin H-J** (2018) Revisiting the Role of Plant Transcription Factors in the Battle against Abiotic Stress. *Int J Mol Sci* **19**: 1634
- Knorst V, Yates S, Byrne S, Asp T, Widmer F, Studer B, Kölliker R** (2019) First assembly of the gene-space of *Lolium multiflorum* and comparison to other Poaceae genomes. *Grassl Sci* **65**: 125–134
- Kuang L, Shen Q, Chen L, Ye L, Yan T, Chen Z-H, Waugh R, Li Q, Huang L, Cai S, et al** (2022) The genome and gene editing system of sea barleygrass provide a novel platform for cereal domestication and stress tolerance studies. *Plant Commun* **3**: 100333
- Kumar V, Khare T, Shaikh S, Wani SH** (2018) Compatible Solutes and Abiotic Stress Tolerance in Plants. *Metab. Adapt. Plants Abiotic Stress*
- Li S** (2015) The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. *Plant Signal. Behav.*
- Li W, Pang S, Lu Z, Jin B** (2020) Function and Mechanism of WRKY Transcription Factors in Abiotic Stress Responses of Plants. *Plants* **9**: 1515
- Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al** (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**: 438–444
- Lundgren MR, Besnard G, Ripley BS, Lehmann CER, Chatelet DS, Kynast RG, Namaganda M, Vorontsova MS, Hall RC, Elia J, et al** (2015) Photosynthetic innovation broadens the niche within a single species. *Ecol Lett* **18**: 1021–1029

- Ma C, Wang H, Macnish AJ, Estrada-Melo AC, Lin J, Chang Y, Reid MS, Jiang C-Z** (2015) Transcriptomic analysis reveals numerous diverse protein kinases and transcription factors involved in desiccation tolerance in the resurrection plant *Myrothamnus flabellifolia*. *Hortic Res.* doi: 10.1038/hortres.2015.34
- Ma Z, Hu L, Jiang W** (2024) Understanding AP2/ERF Transcription Factor Responses and Tolerance to Various Abiotic Stresses in Plants: A Comprehensive Review. *Int J Mol Sci* **25**: 893
- Mamidi S, Healey A, Huang P, Grimwood J, Jenkins J, Barry K, Sreedasyam A, Shu S, Lovell JT, Feldman M, et al** (2020) A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. *Nat Biotechnol* **38**: 1203–1210
- Marks RA, Farrant JM, Nicholas McLetchie D, VanBuren R** (2021a) Unexplored dimensions of variability in vegetative desiccation tolerance. *Am J Bot* **108**: 346–358
- Marks RA, Hotaling S, Frandsen PB, VanBuren R** (2021b) Representation and participation across 20 years of plant genome sequencing. *Nat Plants* **7**: 1571–1578
- Marks RA, Pas LVD, Schuster J, Gilman IS, VanBuren R** (2024) Convergent evolution of desiccation tolerance in grasses. 2023.11.29.569285
- Marques DN, Reis SP dos, de Souza CRB** (2017) Plant NAC transcription factors responsive to abiotic stresses. *Plant Gene* **11**: 170–179
- Minocha R, Majumdar R, Minocha SC** (2014) Polyamines and abiotic stress in plants: a complex relationship<sup>1</sup>. *Front Plant Sci.* doi: 10.3389/fpls.2014.00175
- Mitros T, Session AM, James BT, Wu GA, Belaffif MB, Clark LV, Shu S, Dong H, Barling A, Holmes JR, et al** (2020) Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun* **11**: 5442
- Mondal TK, Rawal HC, Chowrasia S, Varshney D, Panda AK, Mazumdar A, Kaur H, Gaikwad K, Sharma TR, Singh NK** (2018) Draft genome sequence of first monocot-halophytic species *Oryza coarctata* reveals stress-specific genes. *Sci Rep* **8**: 13698
- Muhammad I, Shalmani A, Ali M, Yang Q-H, Ahmad H, Li FB** (2021) Mechanisms Regulating the Dynamics of Photosynthesis Under Abiotic Stresses. *Front Plant Sci.* doi: 10.3389/fpls.2020.615942
- Oh D-H, Kowalski KP, Quach QN, Wijesinghege C, Tanford P, Dassanayake M, Clay K** (2022) Novel genome characteristics contribute to the invasiveness of *Phragmites australis* (common reed). *Mol Ecol* **31**: 1142–1159
- Oliver MJ, Tuba Z, Mishler BD** (2000) The evolution of vegetative desiccation tolerance in

land plants. *Plant Ecol* **151**: 85–100

**Oshunsanya SO, Nwosu NJ, Li Y** (2019) Abiotic Stress in Agricultural Crops Under Climatic Conditions. *In* MK Jhariya, A Banerjee, RS Meena, DK Yadav, eds, *Sustain. Agric. For. Environ. Manag.* Springer, Singapore, pp 71–100

**Pardo J, Man Wai C, Chay H, Madden CF, Hilhorst HWM, Farrant JM, VanBuren R** (2020) Intertwined signatures of desiccation and drought tolerance in grasses. *Proc Natl Acad Sci* **117**: 10079–10088

**Pardo J, VanBuren R** (2021) Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *Plant Cell* **33**: 3391–3401

**Paril J, Pandey G, Barnett EM, Rane RV, Court L, Walsh T, Fournier-Level A** (2022) Rounding up the annual ryegrass genome: High-quality reference genome of *Lolium rigidum*. *Front. Genet.* 13:

**Robbins MD, Bushman BS, Huff DR, Benson CW, Warnke SE, Maughan CA, Jellen EN, Johnson PG, Maughan PJ** (2023) Chromosome-Scale Genome Assembly and Annotation of Allotetraploid Annual Bluegrass (*Poa annua* L.). *Genome Biol Evol* **15**: evac180

**Roy S, Chakraborty U** (2014) Salt tolerance mechanisms in salt tolerant grasses (STGs) and their prospects in cereal crop improvement. *Bot Stud.* doi: 10.1186/1999-3110-55-31

**Sahu A, Singh R, Verma PK** (2023) Plant BBR/BPC transcription factors: unlocking multilayered regulation in development, stress and immunity. *Planta* **258**: 31

**Schubert M, Humphreys AM, Lindberg CL, Preston JC, Fjellheim S** (2020) To Coldly Go Where No Grass has Gone Before: A Multidisciplinary Review of Cold Adaptation in Poaceae. *Annu. Plant Rev. Online.* John Wiley & Sons, Ltd, pp 523–562

**Shi C, Li W, Zhang Q-J, Zhang Y, Tong Y, Li K, Liu Y-L, Gao L-Z** (2020) The draft genome sequence of an upland wild rice species, *Oryza granulata*. *Sci Data* **7**: 131

**Singh J, Thakur JK** (2018) Photosynthesis and Abiotic Stress in Plants. *In* S Vats, ed, *Biot. Abiotic Stress Toler. Plants.* Springer, Singapore, pp 27–46

**Song B, Buckler ES, Wang H, Wu Y, Rees E, Kellogg EA, Gates DJ, Khaipho-Burch M, Bradbury PJ, Ross-Ibarra J, et al** (2021) Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Res* **31**: 1245–1257

**Soreng RJ, Peterson PM, Zuloaga FO, Romaschenko K, Clark LG, Teisher JK, Gillespie LJ, Barberá P, Welker CAD, Kellogg EA, et al** (2022) A worldwide phylogenetic classification of the Poaceae (Gramineae) III: An update. *J Syst Evol* **60**: 476–521

- Strömberg CAE** (2011) Evolution of Grasses and Grassland Ecosystems. *Annu Rev Earth Planet Sci* **39**: 517–544
- Sun G, Wase N, Shu S, Jenkins J, Zhou B, Torres-Rodríguez JV, Chen C, Sandor L, Plott C, Yoshinga Y, et al** (2022) Genome of *Paspalum vaginatum* and the role of trehalose mediated autophagy in increasing maize biomass. *Nat Commun* **13**: 7731
- Tanaka H, Hirakawa H, Kosugi S, Nakayama S, Ono A, Watanabe A, Hashiguchi M, Gondo T, Ishigaki G, Muguerza M, et al** (2016) Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Res* **23**: 171–180
- Tebele SM, Marks RA, Farrant JM** (2021) Two Decades of Desiccation Biology: A Systematic Review of the Best Studied Angiosperm Resurrection Plants. *Plants* **10**: 2784
- Tewari S, Mishra A** (2018) Chapter 18 - Flooding Stress in Plants and Approaches to Overcome. *In* P Ahmad, MA Ahanger, VP Singh, DK Tripathi, P Alam, MN Alyemini, eds, *Plant Metab. Regul. Environ. Stress*. Academic Press, pp 355–366
- Tyagi A, Ali S, Ramakrishna G, Singh A, Park S, Mahmoudi H, Bae H** (2023) Revisiting the Role of Polyamines in Plant Growth and Abiotic Stress Resilience: Mechanisms, Crosstalk, and Future Perspectives. *J Plant Growth Regul* **42**: 5074–5098
- VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, et al** (2020) Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat Commun* **11**: 884
- VanBuren R, Wai CM, Keilwagen J, Pardo J** (2018) A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* **2**: e00096
- Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu J-K** (2006) Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J* **45**: 523–539
- Wang J, Song L, Gong X, Xu J, Li M** (2020) Functions of Jasmonic Acid in Plant Regulation and Response to Abiotic Stress. *Int J Mol Sci* **21**: 1446
- Wang X, Niu Y, Zheng Y** (2021) Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *Int J Mol Sci* **22**: 6125
- Wang Z, Zhu Y, Wang L, Liu X, Liu Y, Phillips J, Deng X** (2009) A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (*BhGolS1*) promoter. *Planta* **230**: 1155–1166
- Yan J, Liu Y, Yang L, He H, Huang Y, Fang L, Scheller HV, Jiang M, Zhang A** (2021) Cell wall  $\beta$ -1,4-galactan regulated by the BPC1/BPC2-GALS1 module aggravates salt

sensitivity in *Arabidopsis thaliana*. *Mol Plant* **14**: 411–425

**Yan Z, Liu H, Chen Y, Sun J, Ma L, Wang A, Miao F, Cong L, Song H, Yin X, et al** (2022) High-quality chromosome-scale de novo assembly of the *Paspalum notatum* ‘Flugge’ genome. *BMC Genomics* **23**: 293

**Yoon Y, Seo DH, Shin H, Kim HJ, Kim CM, Jang G** (2020) The Role of Stress-Responsive Transcription Factors in Modulating Abiotic Stress Tolerance in Plants. *Agronomy* **10**: 788

**Zaid A, Mushtaq M, Wani SH** (2021) Chapter 9 - Interactions of phytohormones with abiotic stress factors under changing climate. *In* T Aftab, KR Hakeem, eds, *Front. Plant-Soil Interact*. Academic Press, pp 221–236

**de Zelicourt A, Colcombet J, Hirt H** (2016) The Role of MAPK Modules and ABA during Abiotic Stress Signaling. *Trends Plant Sci* **21**: 677–685

**Zhang H, Hall N, Goertzen LR, Bi B, Chen CY, Peatman E, Lowe EK, Patel J, McElroy JS** (2019) Development of a goosegrass (*Eleusine indica*) draft genome and application to weed science research. *Pest Manag Sci* **75**: 2776–2784

**Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, et al** (2018) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat Genet* **50**: 1565–1573

**Zhang Q, Song X, Bartels D** (2016) Enzymes and Metabolites in Carbohydrate Metabolism of Desiccation Tolerant Plants. *Proteomes* **4**: 40

**Zhang X, Ekwealor JTB, Mishler BD, Silva AT, Yu L, Jones AK, Nelson ADL, Oliver MJ** (2024) *Syntrichia ruralis*: emerging model moss genome reveals a conserved and previously unknown regulator of desiccation in flowering plants. *New Phytol*. doi: 10.1111/nph.19620



APPENDIX

Table A1.1: Species with sequenced genomes in the grasses by subfamily and tribe. There are 114 total sequenced species in the family (0.97% of total grass species). The most highly sequenced subfamilies are Panicoideae and Pooideae, while the most highly sequenced tribe is Oryzeae.

Subfamily	Tribe	Number of genomes	Total genomes in subfamily
Anomochloideae	Streptochaeteae	1	1
Arundinoideae	Molinieae	1	1
Bambusoideae	Arundinarieae	1	7
	Bambuseae	3	
	Olyreae	3	
Chloridoideae	Cynodonteae	7	17
	Eragrostideae	4	
	Zoysieae	6	
Oryzoideae	Oryzeae	21	21
Panicoideae	Andropogoneae	19	33
	Paniceae	12	
	Paspaleae	2	
Pharoideae	Phareae	1	1
Pooideae	Brachypodieae	3	33
	Poeae	12	
	Stipeae	1	
	Triticeae	17	

Table A1.2: Taxonomic information and availability of genome assemblies for desiccation tolerant grass species. Grasses identified as desiccation tolerant were collected by (Marks et al., 2021a). If there is no reference listed, the genome/transcriptome of that species has not yet been published.

<b>Subfamily</b>	<b>Tribe</b>	<b>Species</b>	<b>Genome reference</b>
Chloridoideae	Cynodonteae	<i>Micrachne patentiflora</i> (also <i>Brachyachne patentiflora</i> )	
Chloridoideae	Cynodonteae	<i>Microchloa caffra</i>	(Marks et al., 2024)
Chloridoideae	Cynodonteae	<i>Microchloa indica</i>	
Chloridoideae	Cynodonteae	<i>Microchloa kunthii</i>	
Chloridoideae	Cynodonteae	<i>Oropetium capense</i>	(Marks et al., 2024)
Chloridoideae	Cynodonteae	<i>Oropetium roxburghianum</i>	
Chloridoideae	Cynodonteae	<i>Oropetium thomaeum</i>	(VanBuren et al., 2018)
Chloridoideae	Cynodonteae	<i>Tripogon capillatus</i>	
Chloridoideae	Cynodonteae	<i>Tripogon curvatus</i>	
Chloridoideae	Cynodonteae	<i>Tripogon filiformis</i>	
Chloridoideae	Cynodonteae	<i>Tripogon jacquemontii</i>	
Chloridoideae	Cynodonteae	<i>Tripogon lisboae</i>	
Chloridoideae	Cynodonteae	<i>Tripogon loliiformis</i>	
Chloridoideae	Cynodonteae	<i>Tripogon major</i>	
Chloridoideae	Cynodonteae	<i>Tripogon minimus</i>	(Marks et al., 2024)
Chloridoideae	Cynodonteae	<i>Tripogon polyanthus</i>	
Chloridoideae	Cynodonteae	<i>Tripogon spicatus</i>	
Chloridoideae	Eragrostideae	<i>Eragrostiella bifaria</i>	
Chloridoideae	Eragrostideae	<i>Eragrostiella brachyphylla</i>	
Chloridoideae	Eragrostideae	<i>Eragrostiella nardioides</i>	
Chloridoideae	Eragrostideae	<i>Eragrostis hispida</i>	
Chloridoideae	Eragrostideae	<i>Eragrostis invalida</i>	
Chloridoideae	Eragrostideae	<i>Eragrostis nindensis</i>	(Pardo et al., 2020)
Chloridoideae	Eragrostideae	<i>Eragrostis paradoxa</i>	

Table 1.2 (cont'd)

Chloridoideae	Zoysieae	<i>Sporobolus atrovirens</i>	
Chloridoideae	Zoysieae	<i>Sporobolus elongatus</i>	
Chloridoideae	Zoysieae	<i>Sporobolus festivus</i>	
Chloridoideae	Zoysieae	<i>Sporobolus fimbriatus</i>	
Chloridoideae	Zoysieae	<i>Sporobolus lampranthus</i>	
Chloridoideae	Zoysieae	<i>Sporobolus pellucidus</i>	
Chloridoideae	Zoysieae	<i>Sporobolus stapfianus</i>	(Chávez Montes et al., 2022)
Micrairoideae	Micraireae	<i>Micraira adamsii</i>	
Micrairoideae	Micraireae	<i>Micraira lazaridis</i>	
Micrairoideae	Micraireae	<i>Micraira multinerva</i>	
Micrairoideae	Micraireae	<i>Micraira spinifera</i>	
Micrairoideae	Micraireae	<i>Micraira subulifolia</i>	
Micrairoideae	Micraireae	<i>Micraira tenuis</i>	
Micrairoideae	Micraireae	<i>Micraira viscidula</i>	
Pooideae	Poeae	<i>Poa bulbosa</i>	
Pooideae	Poeae	<i>Poa eigii</i>	

Table A1.3: Sequenced grass species identified as salt tolerant by (Bennett et al., 2013). The highest number of sequenced salt tolerant species was found in the Panicoideae.

<b>Subfamily</b>	<b>Tribe</b>	<b>Species</b>	<b>Genome Reference</b>
Arundinoideae	Molinieae	<i>Phragmites australis</i>	(Oh et al., 2022)
Chloridoideae	Cynodonteae	<i>Eleusine indica</i>	(Zhang et al., 2019)
Chloridoideae	Eragrostideae	<i>Eragrostis curvula</i>	(Carballo et al., 2019)
Chloridoideae	Eragrostideae	<i>Eragrostis pilosa</i>	
Chloridoideae	Zoysieae	<i>Sporobolus alterniflorus</i> (formerly <i>Spartina alterniflora</i> )	(Chen et al., 2024)
Chloridoideae	Zoysieae	<i>Zoysia japonica</i>	(Tanaka et al., 2016)
Chloridoideae	Zoysieae	<i>Zoysia matrella</i>	(Tanaka et al., 2016)
Oryzoideae	Oryzeae	<i>Oryza coarctata</i>	(Mondal et al., 2018)
Oryzoideae	Oryzeae	<i>Oryza meyeriana</i> (sequenced: var. <i>granulata</i> )	(Shi et al., 2020)
Panicoideae	Andropogoneae	<i>Chrysopogon serrulatus</i>	(Song et al., 2021)
Panicoideae	Andropogoneae	<i>Coix lacryma-jobi</i>	(Kang et al., 2020)
Panicoideae	Andropogoneae	<i>Cymbopogon citratus</i>	(Chakravartty et al., 2023)
Panicoideae	Andropogoneae	<i>Cymbopogon flexuosus</i>	
Panicoideae	Andropogoneae	<i>Miscanthus sinensis</i>	(Mitros et al., 2020)
Panicoideae	Andropogoneae	<i>Saccharum spontaneum</i>	(Zhang et al., 2018)
Panicoideae	Paniceae	<i>Echinochloa crus-galli</i>	(Guo et al., 2017)
Panicoideae	Paniceae	<i>Setaria viridis</i>	(Mamidi et al., 2020)
Panicoideae	Paspaleae	<i>Paspalum notatum</i>	(Yan et al., 2022b)
Panicoideae	Paspaleae	<i>Paspalum vaginatum</i>	(Sun et al., 2022)
Pooideae	Poeae	<i>Alopecurus myosuroides</i>	(Cai et al., 2023)
Pooideae	Poeae	<i>Lolium multiflorum</i>	(Knorst et al., 2019)
Pooideae	Poeae	<i>Lolium rigidum</i>	(Paril et al., 2022)
Pooideae	Poeae	<i>Poa annua</i>	(Robbins et al., 2023)
Pooideae	Triticeae	<i>Hordeum marinum</i>	(Kuang et al., 2022)
Pooideae	Triticeae	<i>Hordeum vulgare</i>	(Beier et al., 2017)

## **Chapter 2: Stress-responsive transcription factor families are key components of the core stress response in maize**

### **ABSTRACT**

Abiotic stresses, including drought, salt, heat, cold, flooding, and low nitrogen among others, can have devastating impacts on agriculture and are increasing in frequency worldwide due to climate change. Maize is one of the most important food and feed crops globally and may experience multiple abiotic stresses of different types throughout its growing season. Thus, improving the stress tolerance of maize is a top priority for agricultural resilience. Plants share many physiological and transcriptomic responses to different abiotic stresses, many of which are regulated by transcription factors in stress-responsive families. We identified a set of core abiotic stress genes in maize by meta-analysis of nearly 1,900 RNA-sequencing samples and found that the core genes are enriched in stress-responsive transcription factors, including specific families such as AP2/ERF-ERF, NAC, bZIP, HSF, and C2C2-CO-like. Co-expression network analysis revealed that core stress transcription factors of these enriched families are always found in the same modules as at least two sets of stress-specific genes, indicating that these transcription factors are putative regulators not only of the core stress response, but of stress-specific responses as well.

### **INTRODUCTION**

Abiotic stresses such as drought, salt, flooding, heat, cold, and low nitrogen can have severe impacts on agricultural crops. Climate change has slowed the growth of agricultural productivity in the last half century, and events such as droughts and floods are projected to increase in frequency with changing climate (Calvin et al., 2023). This is expected to cause a general decrease in global crop productivity (Raza et al., 2019). Maize (*Zea mays*) is one of the most important food, feed, and biofuel crops globally, and like all crops, may experience multiple abiotic stressors individually or in combination throughout the growing season. Thus, understanding the basis of abiotic stress response on the molecular level is important for eventual improvement of maize stress tolerance and increased agricultural resilience in the face of climate change.

Abiotic stresses present many challenges to plants at the cellular and molecular levels. For example, most abiotic stressors including cold, low nitrogen, drought, and flooding are known to inhibit photosynthesis (Singh and Thakur, 2018; Li et al., 2019; Wu et al., 2019; Qi et

al., 2021). Typically this leads to a buildup of reactive oxygen species, which can damage cell components such as membranes. Antioxidants such as glutathione, which scavenge excess reactive oxygen species, are thus instrumental in the cellular response to stresses including salt, drought, flooding, and cold (Aslam et al., 2021). In stresses with an osmotic component, such as drought, salt, and cold, it is common for plants to accumulate osmolytes/compatible solutes such as sugars, amino acids including proline and glycine betaine, and polyamines; these molecules function as protectants against water loss and may also have antioxidant activity (Sharma et al., 2019). Polyamines are also hypothesized to act as plant growth regulators (EL Sabagh et al., 2022), and both overexpression of polyamine biosynthetic genes and exogenous polyamine application have been shown to improve plant stress tolerance (Minocha et al., 2014). Many of these responses are mediated by phytohormones such as ethylene, abscisic acid (ABA), and jasmonic acid (JA), and are similar across different stressors. For example, ethylene and JA are noted as active regulators under various stress conditions, including high and low temperature, drought, and salt among others (EL Sabagh et al., 2022).

The stress responses discussed above are generally regulated by transcription factors, proteins that alter the expression levels of stress-responsive genes; depending upon the TF family, gene regulation may occur in a phytohormone-dependent or independent manner. For example, TFs in the ABF subfamily of the bZIP family are activated by signal transduction following ABA recognition; following their activation, they go on to activate the expression of NAC and AP2/ERF family TFs (Yoon et al., 2020). These TFs then go on to regulate other stress-responsive genes (Mizoi et al., 2012; Nakashima et al., 2012; Shao et al., 2015).

Transcriptomics is a common approach when studying abiotic stress response, and studies of core stress responses are no exception. Generally, core stress responses have been studied using two approaches. The first involves gathering transcriptomic data from an experiment with multiple different stress treatments and the second involves meta-analysis of previously published transcriptomic data, either from microarrays or RNA-sequencing. In both cases, differentially expressed genes (DEGs) are usually calculated between stressed and control conditions, and the DEGs from different stresses are then overlapped in an approach we term “set operations” to find the core genes. Meta-analyses and multi-stressor experiments have previously been conducted in cotton (Tahmasebi et al., 2019), rice (Cohen and Leach, 2019), sesame (Dossa et al., 2019), *Brassica napus* (Zhang et al., 2019), and *Arabidopsis thaliana*

(Sanchez-Munoz et al., 2024; Shintani et al., 2024), as well as maize (Li et al., 2017), among others. These previous meta-analyses often use a relatively limited number of studies and sometimes only one per stressor, which may limit the power of the meta-analysis to draw biologically relevant conclusions. All meta-analyses cited here re-analyzed data from at most five hundred transcriptome samples. The largest of these, by (Sanchez-Munoz et al., 2024), re-analyzed 500 samples from 23 different studies, including both microarray and RNA-seq data, from a total of 11 stressors in both roots and shoots. Most other meta-analyses examined only four or five stressors.

In the current study, we re-analyzed nearly 1,900 RNA-sequencing samples from 39 different maize stress experiments, spanning six abiotic stresses as well as a wide variety of genotypes, growth environments, tissues, and developmental stages. We leveraged these data to find core genes via both the standard set operations approach and random forest classification, which is a novel machine learning approach for the core stress response field. As part of our set operations, we also identified peripheral, or stress-specific, genes. We identified core and peripheral genes both for all tissues in the dataset and for photosynthetic tissues only, due to the variability in gene expression between photosynthetic and non-photosynthetic tissues. Analysis of the core genes revealed that they were enriched in several stress-related TF families, which were found to be co-expressed both with other core genes and with peripheral genes, indicating their possible role in regulating not only the core response, but stress-specific responses as well.

## METHODS

### *Curating maize RNA-seq data*

This study utilized publicly available and previously published abiotic stress RNA-sequencing (RNA-seq) data in maize from the NCBI Sequence Read Archive (SRA). Only data that could be linked with published papers was used, and data was gathered for drought, cold, heat, salt, flooding, and low nitrogen. Polyethylene glycol and similar osmotic stress treatments such as sugar alcohols were not included in the drought data. These six abiotic stressors are the most commonly studied in maize, and each had at least three independent experiments (BioProjects) with high quality data. BioProjects were only included in the study if they had at least one well-documented stress time point as well as either a control treatment or samples taken at experiment initiation. Thirty-nine BioProjects containing 1,872 samples total met these criteria, and were selected for use in this study. The dataset includes stress-tolerant and sensitive

maize genotypes, and hybrids, but not mutants or transgenic plants. Most samples were from leaf or other photosynthetic tissues, but roots, reproductive, and seed tissues were also included (Table A2.1). Studies with any number of replicates were used.

#### *Processing the RNA-seq data*

All data were downloaded from the SRA using the `prefetch` and `fasterq-dump` commands from `sra-toolkit` version 2.11.2 (<https://github.com/ncbi/sra-tools>). The raw RNA-seq reads were processed using the Nextflow `nf-core rnaseq` pipeline (Di Tommaso et al., 2017; Ewels et al., 2020; dependencies: <https://github.com/nf-core/rnaseq/blob/master/CITATIONS.md>). Reads were trimmed using `fastp` (Chen et al., 2018) and transcripts were quantified using Salmon (Patro et al., 2017). Length-scaled transcripts per million (TPM) were then generated using `tximport` (Soneson et al., 2016). The exception to our processing workflow was a low nitrogen dataset (Ying et al., 2023; BioProject: PRJNA904734) procured directly from one of the authors, which had already been processed with STAR (Dobin et al., 2013). The `convertCounts()` function in the `DGEobj.utils` R package v1.0.6 (<https://CRAN.R-project.org/package=DGEobj.utils>) was used in R 4.2.1 (R Core Team, 2022) to convert the provided count matrix into TPM.

All transcripts were pseudoaligned to the B73 v5 genome (Hufford et al., 2021). Our dataset contains diverse maize genotypes, including multiple with chromosome scale assemblies, and we initially processed the data by pseudoaligning each genotype to the corresponding reference genome when available. However, we ultimately decided to pseudoalign all data to B73 for the following reasons: 1) mapping rates were similar when pseudoaligning to B73 vs. the corresponding genome for several inbred lines (see Table B2.1); 2) many genotypes do not have sequenced genomes or are hybrids; 3) the B73 reference annotation is manually curated and more complete than others; and 4) graph based annotations and syntenic gene groupings are still incomplete, making it challenging to create an informative set of comparable genes across genotypes for downstream comparisons. Thus, given the goals of this study, and the above limitations, we chose to use a single genome for read mapping.

#### *Data exploration of experimental factors*

Data exploration was conducted with principal component analysis (PCA) using the full dataset that includes 12 tissue types (“all tissues” hereafter), and for photosynthetic tissues only, which included leaf, leaf meristem, and shoot samples. The raw gene expression (TPM) values were filtered to remove genes with zero variance across samples, and subsequently  $\log_2$



transformed after adding 1 to each value using numpy v1.24.3 (Harris et al., 2020). The gene expression data was collected using different sequencing machines, read lengths, coverage, and under different experimental conditions, and we reduced the batch effect of BioProject using pyComBat v0.3.3 (Behdenna et al., 2023). PCA was run on the uncorrected and batch corrected data for the full dataset (all tissues) and photosynthetic data only, and the first two principal components were calculated using scikit-learn v1.2.2 (Pedregosa et al., 2011) and plotted using matplotlib v3.7.1 (Hunter, 2007).

### *Linear modeling*

To further explore heterogeneity within the dataset, we modeled the first principal component of log-transformed and non-batch corrected TPM as a function of genotype, BioProject, treatment, and tissue, using the `lm()` function in R v4.2.1 (R Core Team, 2022). No interaction effects were included for computational efficiency, and due to the difficulty of biological interpretation of interaction effects' meanings. This modeling approach was repeated for PC1 of batch corrected TPM, allowing us to compare results between the two models.

### *Differential expression analysis and identification of stress-induced genes*

Changes in gene expression between stressed and control samples were evaluated using two different methodologies: differential expression analysis and fold change or TN-ratio (described in detail below) (Shintani et al., 2024). For differential expression analysis, DESeq2 (Love et al., 2014) was used to identify statistically differentially expressed genes. Differential expression was calculated as a function of an identifier containing genotype, treatment, time point, developmental stage, and tissue information for each sample. The default model was used. Only sets of samples containing at least 3 replicates were used for differential expression.

The TN-ratio is the ratio of transcripts per million (TPM) between stress-treated (T) and control, or non-treated (N) samples (Shintani et al., 2024). TN-ratio was calculated using the formula from (Shintani et al., 2024) as follows:

$$\text{TN-ratio} = (\text{stress-treated TPM} + 1) / (\text{non-treated TPM} + 1)$$

For our study, TN-ratio was calculated on a per-experiment basis where for a given BioProject the mean TPM was calculated for each treatment of control or stress treated, and this was used to calculate TN-ratio.

We used a similar criteria as outlined in (Shintani et al., 2024), where genes with a TN-ratio of greater than 2 were considered upregulated, and those with a TN-ratio of less than 0.5

were considered downregulated. For set operations, the union of upregulated and downregulated genes for each experiment was calculated and the set of core genes among all six stressors was defined as the six-way intersection of individual sets. A set of core genes was identified using all samples in the dataset and using the samples from only photosynthetic tissues (leaf, leaf meristem, and shoot). Genes that were only differentially expressed for a subset or only one stress condition were also identified, and we refer to these sets as ‘peripheral genes’ or stress-specific genes. The overlap of up- and downregulated genes among experiments within each stressor was also examined.

#### *Hierarchical clustering of abiotic stresses*

To determine the relationships of transcriptomic responses to different abiotic stresses, we performed hierarchical clustering. BioProject-corrected, log<sub>2</sub> transformed TPM values were scaled to a z-score using scikit-learn and used as input for both the full sample set and photosynthetic tissues only. For each of the seven treatments (six stress conditions and control), a mean expression value of the scaled and transformed TPM data was calculated and used as input for hierarchical clustering and dendrogram visualization using scipy v1.10.1 (Virtanen et al., 2020).

#### *Random forest binary classification*

Random forest models were used to classify whether samples were stressed or control. To avoid data leakage, all stressed and associated control samples for a single stressor were held out for use as the test set. Given the hypothesized existence of a core stress response transcriptome, a random forest model tested on a stress it was not trained on was hypothesized to be able to accurately classify stressed and control samples. This was repeated for all stressors, so that each stressor was used as the test set once, resulting in a total of six models, each with separately tuned hyperparameters. Hyperparameters tuned for each model included bootstrapping, maximum tree depth, maximum features, minimum number of samples per leaf, minimum samples split, and number of estimators. In each iteration, all other samples were used for the training set. This modeling strategy was applied to both all tissues (full dataset) and samples from photosynthetic tissues only.

The BioProject-corrected and log<sub>2</sub>-transformed TPM were used as features in the model such that each feature was a maize gene. SMOTE was used for upsampling to balance numbers of control vs. stress using training data. As stated above, hyperparameters were tuned separately

for each model, with individual models having different stressors as the test set. The optimal hyperparameters were then used for training and making predictions.

For each of the six models fit for each set of samples, feature importance was calculated for all features (genes) used in the model. For each model, evaluation of possible core gene sets was conducted via iterative feature selection, as follows. The corrected TPM were filtered to only the top X features, where X=50, 100, 250, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 10,000, or 15,000. Following TPM subsetting, a new RF model was run on each subset and the model performance metrics accuracy, AUC, and F1 were calculated for each model. Based on the optimum model performance using the smallest subset of features, the top 6,000 most important features were extracted from each of the six individual stressor models. The intersection of these six sets of 6,000 genes each was calculated to get the core stress genes from random forest for each set of samples (all tissues and photosynthetic tissues).

#### *Co-expression network analysis*

A co-expression network was constructed with the corrected TPM of the full maize dataset using Weighted Gene Coexpression Network Analysis (WGCNA) (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012). A soft threshold of 9 was used for network construction. Hub genes, i.e. genes that had a high positive or negative correlation with most genes in the same module, were identified using the module membership generated with the signedKME() function from WGCNA. Only genes with module membership >0.86 (95th percentile of the absolute values of module membership) were considered hub genes. We then found which hub genes were found in each core gene set, and used Fisher's exact test implemented in Python using scipy.stats as described above to test whether there were more core genes in the set of hub genes than expected by chance.

#### *GO term enrichment*

Gene Ontology (GO) term enrichment was performed with topGO v2.50.0 (Alexa and Rahnenfuhrer, 2022) separately for Biological Process GO terms in upregulated and downregulated core genes from the set operations, random forest, and combined approaches, for both sets of samples. We also ran GO enrichment for the peripheral stress genes for each stressor, separately for upregulated and downregulated peripheral sets, for each set of tissues, and for the genes in each co-expression module. Fisher's exact test was used as the enrichment test, and the classic algorithm was used. False discovery rate (FDR) p-value correction was used to

adjust p-values, and an FDR adjusted p-value of less than 0.05 was considered statistically significant. GO terms were current as of March 25, 2024.

### *Transcription factor enrichment*

The list of transcription factors (TFs) in the *Z. mays* B73 v5 genome was downloaded from Grassius (<https://grassius.org/species/Maize>, Yilmaz et al., 2009). We tested for enrichment of TFs from all families (not any particular family) in the upregulated and downregulated core genes using a one-sided Fisher's exact test implemented using `scipy.stats` v1.10.1 (Virtanen et al., 2020). We also tested for TF enrichment in the upregulated and downregulated peripheral genes from each stressor. The “`fdr correction`” function from `statsmodels.stats.multitest` v0.13.5 (Seabold and Perktold, 2010) was used to adjust the resulting p-values. As during GO enrichment, FDR-adjusted P values less than 0.05 were considered significant.

Again using information from Grassius, the sets of all upregulated core genes and all downregulated core genes for each tissue set, from both methods combined, were tested for enrichment of each TF family. TF families with FDR-adjusted P values of less than 0.05 were considered significantly enriched, while those with adjusted P values of between 0.1 and 0.05 were considered close to enriched and thus also selected for further analysis.

For the subset of core genes belonging to the enriched or near-enriched TF families, we identified what co-expression network module each gene belonged to. We then used Fisher's exact test via `scipy.stats` as above to determine whether modules containing these transcription factors of interest were enriched in core genes. Again, FDR was used to adjust P values and adjusted P of less than 0.05 was considered significant.

## RESULTS

### *Data exploration*

To search for conserved molecular signatures of abiotic stress responses, we gathered published maize RNA-seq data from the NCBI Sequence Read Archive. A total of 39 different BioProjects were selected for use in this study. This includes 15 for drought, 8 for heat, 8 for cold, 5 each for low nitrogen and salt, and 3 for flooding (Figure A2.1A). There were 1,872 total samples in the dataset; drought had the most samples, while salt had the fewest (Figure A2.1B). The most prevalent tissue type was leaves, followed by roots; various other vegetative and reproductive tissues were also included (Table A2.1). Samples were collected in greenhouse, field, and growth chamber environments with developmental stages ranging from germination to

reproduction. Expression data spanned 328 different maize genotypes, including genotypes listed as both sensitive and tolerant to various stress conditions. Inbred lines B73, W22, and Mo17 were the most common genotypes used across all expression experiments (Figure A2.1D). There was also intra-treatment heterogeneity across studies, as illustrated by the range of temperatures used as cold, heat, and control conditions; the high end of the cold-treatment temperatures almost overlaps with the low end of control temperatures, and vice versa for heat (Figure A2.1C).

All RNA-seq samples were downloaded from the SRA and processed using the Nextflow nf-core rnaseq pipeline (Di Tommaso et al., 2017; Ewels et al., 2020; dependencies: <https://github.com/nf-core/rnaseq/blob/master/CITATIONS.md>). The B73 maize reference genome has the most complete annotation and was used for quantifying expression of all genotypes. Although substantial variation in gene content has been observed across maize diversity, a single reference enables comparisons across all datasets. Briefly, samples were trimmed using fastp and pseudoaligned to the B73 V5 genome using Salmon. Most reads pseudoaligned to the B73 genome with a mapping rate of greater than 60% (Figure B2.1). We then used tximport to generate length scaled transcripts per million (TPM) which were used for downstream analyses.

These TPM were then log-transformed and fed into a principal component analysis to explore how the data would separate. Approximately 28% of the variance in the full dataset was explained by PC1 and 15% by PC2. We found that samples grouped first by tissue type, with photosynthetic and non-photosynthetic tissues segregating along PC1 for the most part, then by BioProject (Figure A2.2A and B). There was little apparent clustering of samples by treatment (Figure A2.2D), growth environment (Figure A2.2C), or developmental stage (Figure B2.2). Any clustering by these factors is likely an artifact of BioProject, as the authors of each experiment used different developmental stages and growth environments in their studies, and may have also imposed their stress treatments differently.

We conducted linear modeling on the first principal component of gene expression, with genotype, BioProject, tissue, and treatment as independent variables. The p-value for each of the four independent variables was less than 0.001, indicating that each independent variable had a highly significant effect on gene expression. Thus, we can conclude that heterogeneity in all of these factors significantly contributes to the variability in gene expression among samples. We used pyComBat (Behdenna et al., 2023) to adjust for batch effects and reduce the variance due to

BioProject, and re-ran the PCA with the corrected expression values (Figure A2.2, right column). For the pyComBat adjusted expression, PC1 explained about 16% of variance, and PC2 explained about 14%, and we observed an overall reduction in grouping due to BioProject (Figure A2.2B). However, some grouping by BioProject was still evident, and this may be due to variability in other factors, such as growth environment (Figure A2.2C). We used this pyComBat adjusted expression matrix for all downstream analyses.

#### *Relationships among stress response transcriptomes*

Following basic exploration of the dataset, we examined the relative similarity of transcriptomic responses to different abiotic stresses using hierarchical clustering on corrected TPM, both for all tissues and for photosynthetic tissues only. The clustering results differed depending on the tissue set used (Figure A2.3). For photosynthetic tissues, the hierarchical clustering identified two major clusters, with low nitrogen stress as an outgroup. One cluster contained control, heat, and cold; the other contained drought, salt, and flooding. Two main clusters were also identified for all tissues (Figure A2.3); however, they differed substantially in composition, with flooding and salt comprising one cluster, and the other containing control, cold, low nitrogen, heat, and drought. This difference is likely due to differences in tissue-specific abiotic stress responses; the inclusion of roots and reproductive tissues in the “all tissues” set must substantially change the overall transcriptomes of the various stressors, especially since the low nitrogen gene expression in this dataset mostly comes from roots.

#### *Core stress gene set identification and characterization*

To identify the core stress-responsive genes across multiple abiotic stressors in maize, we used both set operations of the ratio of gene expression under treatment vs normal conditions (fold change or TN-ratios; hereon, set operations) and a machine learning model based on random forest algorithms. Set operations is the typical method used for identification of core genes in prior meta-analyses (Cohen and Leach, 2019; Dossa et al., 2019; Tahmasebi et al., 2019; Zhang et al., 2019; Shintani et al., 2024). We applied random forest binary classification to broaden our search for core genes, thus ensuring the robustness of our study; the most important predictive features (i.e., genes) in the random forest were considered core genes (see Methods). This is a novel method in the core stress literature, although support vector machine clustering, another machine learning method, was previously used in (Sanchez-Munoz et al., 2024). However, our random forest approach differs from the previous machine learning approach in

that it utilizes classification rather than clustering. Given our underlying hypothesis that a core stress transcriptome exists, we expected that a binary random forest classifier model would be able to predict whether a given transcriptome was from a stressed or control sample, even if the model had not been trained on the stressor on which it was being tested. This led to our “hold one stressor out” random forest approach (see Methods for more details).

The efficacy of random forest prediction varied across stressors and, to a lesser extent, tissue sets (Figures B2.3-B2.4), although all area under the ROC curve (AUC) values were greater than 0.5. Salt and drought were consistently predicted most accurately, while low nitrogen, cold, and heat all had AUC values of slightly greater than 0.5. This is similar to the close clustering of temperature stressors with control after hierarchical clustering of treatments in both all tissues and photosynthetic tissues (Figure A2.3), which can further be explained by the fact that the low end of control temperatures overlaps with the high end of cold-treatment temperatures, and vice versa for heat (Figure A2.1C). Thus, some “heat” and “cold”-treated plants may not have been fully physiologically stressed. These results indicate that our random forest approach was able to pick up on biology related to stress response, particularly core stress response; thus, the random forest top features are useful for defining core genes.

Using these combined methods, we identified core gene sets for all tissues and a subset of photosynthetic tissues only. As shown by PCA (Figure A2.2A), tissue is the most important grouping factor for gene expression in our dataset; most prominently, photosynthetic tissues separate from non-photosynthetic tissues along PC1. Further, most prior core stress meta-analyses have not differentiated between core gene sets for different tissues, with the exception of (Sanchez-Munoz et al., 2024). Table A2.2 shows a summary of different categories of the core gene sets; notably, in all cases more core genes were found via RF than by set operations, and more core genes are upregulated than downregulated under stress conditions. The top features in the random forest models were generally differentially expressed under stress conditions, and for all tissues, there were 11 core genes that were not differentially expressed and for photosynthetic tissues, there were six. In all tissues, a total of 744 core genes were identified, whereas for photosynthetic tissues only there were 512.

Once core genes were identified, we tested for overlaps among these genes across different methods, tissue sets, and whether they were up- or downregulated. We found that genes identified by the same method, whether by random forest or set operations, showed more

similarity across different tissue sets than those identified by different methods within the same tissue set (Figure A2.4A). Core genes identified by random forest may be differentially expressed in opposite directions for individual stresses, and we observed a large overlap between upregulated and downregulated genes (Figure A2.4B and C). This was not the case for core genes from set operations, where by definition the upregulated and downregulated genes were identified separately. There was minimal overlap between core genes identified from different methods for each tissue set (Figure A2.4B and C), suggesting that the random forest core genes, identified based on their importance to the models rather than strictly by differential expression, may represent an emergent core response that would not be identified by set operations. It is our opinion that future core stress meta-analyses could benefit from a similar use of machine learning to identify emergent core stress genes.

To investigate the functions of the core abiotic stress genes in maize, we ran Biological Process GO term enrichment separately on the upregulated and downregulated sets of core genes from each tissue set and method, including both methods combined. No enriched GO terms were found for core genes from photosynthetic tissues; however, several enriched terms were found for the core genes from all tissues (Figure A2.4D). These were largely distinct between upregulated and downregulated core genes, but four terms, “biological regulation”, “response to oxygen-containing compound”, “response to abiotic stimulus”, and “response to temperature stimulus”, were found across up- and downregulated core gene sets, albeit sometimes with different p-values (Figure A2.4D). This makes sense because generic responses to stimuli may take the form of either increases or decreases in abundance, and regulation may be applied by repression or activation.

Terms specific to the downregulated core genes included, notably, various terms related to polyamine metabolism, especially that of the “higher” polyamines spermidine and spermine (Figure A2.4D), which have 3 and 4 amine groups, respectively. Polyamines are known to be stress-involved molecules, in many cases stress-protective; although the molecular mechanism of this protection is not yet known in detail, plants that overexpress polyamine biosynthetic genes and consequently accumulate polyamines often display improved stress tolerance (Minocha et al., 2014; Bano et al., 2020). However, polyamine catabolism can also release reactive oxygen species, so it is possible that polyamines may also contribute to oxidative stress (Minocha et al., 2014). The GO term “circadian rhythm” was also uniquely enriched in downregulated core genes



(Figure A2.4D). Interactions between the circadian clock and abiotic stress responses are complex, but various clock components have been found to be downregulated under different stressors (Sharma et al., 2022).

In upregulated core genes, specifically those found via set operations, other metabolism-related GO terms were enriched: “cellular modified amino acid metabolic process” and “glutathione metabolic process” (Figure A2.4D). Amino acids are integral to stress response, for example, as compatible solutes; proline is particularly noted for behaving in this manner (Batista-Silva et al., 2019). Glutathione is also a noted important antioxidant under various stressors (Aslam et al., 2021). In addition, many terms related to transcriptional regulation (i.e. “regulation of RNA biosynthetic process”) were enriched in the upregulated core genes, from both methods combined. This led us to investigate the presence of transcription factors among the core genes.

#### *Peripheral stress gene set identification and characterization*

The peripheral genes were defined as those that were differentially expressed in response to only one stress condition, in at least one study of that stressor, and were identified by set operations. Similar to core genes, peripheral genes were identified for both all tissues and photosynthetic tissues only. The number of peripheral genes varied by stressor (Table A2.3), with, flooding consistently having the fewest peripheral genes, likely because it also had the fewest BioProjects in the dataset (Figure A2.1A). Across tissue sets, heat had more upregulated than downregulated peripheral genes, while drought and cold had more downregulated peripheral genes (Table A2.3).

Notably, two of the GO terms enriched in upregulated peripheral genes in cold stress in all tissues were related to microtubules (Figure B2.5). Microtubules have been found to act as sensors for various stressors including cold stress (Ma and Liu, 2019). These GO terms, however, were not enriched in core stress genes, indicating that microtubule stress sensing is not a core stress response.

There are eleven GO terms that are enriched both in core genes from all tissues and peripheral genes from all tissues. All of them are related to regulation, including “regulation of DNA-templated transcription”, “regulation of macromolecule biosynthetic process”, and “regulation of biosynthetic process”. These terms were enriched only in downregulated peripheral genes from cold stress (Figure B2.5), while for core genes, they were enriched in

upregulated core genes overall (Figure A2.4D). Thus it is possible that the cold stress response in maize is at least partially regulated by upregulated core stress genes.

#### *Transcription factor enrichment*

We used Fisher's exact test to test for enrichment of transcription factors in general and specific TF families in sets of core genes for both tissues. In all cases, upregulated and downregulated core genes were tested separately; for general TF enrichment tests, core gene sets were also separated by method. For core stress genes from photosynthetic tissues, both upregulated ( $P < 0.001$ ) and downregulated ( $P = 0.0126$ ) core genes from random forest were enriched in TFs in general, as was the total set of upregulated core genes ( $P < 0.001$ ). There were no significantly enriched or near-enriched ( $0.05 < P < 0.1$ ) TF families in the core genes from photosynthetic tissues, except the orphan TFs, which were near-enriched in downregulated core genes ( $P = 0.0847$ ; see Table A2.4).

For the core stress genes from all tissues, general TF enrichment was found to be significant for the upregulated and downregulated genes both from RF and from both methods combined (all  $P < 0.001$ ). However, in this case, significantly enriched and near-enriched TF families were also found (see Table A2.4). Except for orphans, which do not constitute a true TF family but rather a catch-all group for likely TFs and co-regulators that don't clearly fall into another family (<https://grassius.org/family/Maize/Orphans>), the enriched and near-enriched families have all been previously related to stress tolerance. These include the bZIP family, of which the ABF subfamily is directly involved in ABA signaling (Yoon et al., 2020); the ERF subfamily of the AP2-ERF family, which are known to be involved in flooding response (Mizoi et al., 2012); and the HSFs, which regulate heat shock proteins (HSPs) (Andrási et al., 2021), which are important molecular chaperones acting in response to various abiotic stressors (ul Haq et al., 2019).

We also ran TF enrichment for the peripheral stress genes from both tissue sets, treating upregulated and downregulated genes separately as for the core genes. Upregulated peripheral genes for cold stress were enriched in TFs in general both for all tissues ( $P < 0.001$ ) and for photosynthetic tissues ( $P = 0.0087$ ). TFs were also enriched in downregulated peripheral genes for cold stress, only in all tissues ( $P < 0.001$ ). In photosynthetic tissues, upregulated low nitrogen-specific genes were enriched in TFs in general ( $P = 0.0076$ ). For all tissues, the following stress-specific gene sets were enriched in TFs: upregulated flooding-specific ( $P = 0.0422$ ) and

downregulated heat-specific ( $P < 0.001$ ). Table B2.2 shows the TF families that were enriched in different peripheral gene sets. These were largely different from the families enriched in core genes, but there were some families enriched in both core and peripheral sets. For all tissues, these included AP2/ERF-ERF, which was enriched in upregulated core genes and near-enriched in upregulated heat-specific genes; C2C2-CO-like, which was enriched in upregulated core genes and in upregulated flooding-specific genes; and NAC, which was near-enriched in upregulated core genes and enriched in upregulated heat, downregulated drought, and downregulated flooding-specific genes. ERF and NAC TFs, in particular, are noted for their stress responsiveness, so it is not surprising to find them enriched in both core and peripheral gene sets.

#### *Co-expression network analysis*

A co-expression network was generated from the batch corrected TPM for all samples using WGCNA (Langfelder and Horvath, 2008; Langfelder and Horvath, 2012). The co-expression network contained a total of 37,216 genes in 23 modules, with a mean module size of 1,618 genes. The largest module (blue) contained 10,959 genes; the smallest (white) contained 103 genes.

Of the 23 modules, 20 contained at least one core gene from both tissue sets; the darkorange and darkturquoise modules contained no core genes, and the darkgrey module contained no core genes from photosynthetic tissues, but did contain core genes from all tissues. The royalblue module contained the highest percentage of core genes from all tissues, and the lightcyan module contained the highest percentage of core genes from photosynthetic tissues. The five modules with greater than 5% core genes (all tissues) were royalblue, pink, lightgreen, grey60, and green; we ran GO term enrichment for these modules to find out about their functions. Enriched GO terms were non-overlapping for four of these five modules (Figure A2.6C; there were no enriched GO terms found for the green module). GO terms for the lightgreen module pertained only to ethylene response; for the grey60 module, enriched GO terms were largely metabolic, particularly for nitrogen metabolism; the pink module had the most enriched GO terms, many of which were related to regulation; and the royalblue module contained various stimulus and stress response terms, as well as protein folding terms, indicating there may be chaperones present within this module.

13 of the co-expression modules contained TFs previously identified within the core gene set for all tissues, and 12 modules contained TFs from the photosynthetic tissues' core gene set.

Eight modules within each of these groups were also enriched in non-TF core genes at  $p < 0.05$ , Fisher's exact test.

Hub genes were identified from the co-expression data using the module membership correlation values, and genes with module membership above the 95th percentile were considered hub genes. There were 1,861 hub genes total. Of these, 26 were photosynthetic tissues core genes, which is not significantly more than expected by chance (Fisher's exact test,  $p = 0.400$ ). From the full tissue set, 14 hub genes were core genes, which was also not significantly more than expected by chance (Fisher's exact test,  $p = 0.999$ ). None of the core hub genes were transcription factors.

We also tested whether the co-expression modules were enriched in peripheral genes for each stressor and in core genes using Fisher's exact test (Figure A2.6A and B). This was repeated for each set of tissues, as in other analyses. In all modules where core stress genes were enriched, at least some sets of peripheral genes were also enriched. It follows that the core genes are not more co-expressed with each other than with peripheral genes.

The blue, grey60, green, turquoise, lightgreen, royalblue, pink, orange, and greenyellow modules all contained at least one core gene TF of one of the enriched or near-enriched families (Table A2.4), excluding orphans. All these modules except greenyellow and orange were enriched in core genes in all tissues (Figure A2.6A), and all were enriched in at least 2 peripheral gene sets. This indicates that the core gene members of the AP2/ERF-ERF, NAC, bZIP, HSF, and C2C2-CO-like TF families are co-expressed with stress-specific peripheral genes and may regulate these genes.

## DISCUSSION

Abiotic stresses such as drought, flooding, salinity, and temperature extremes trigger similar physiological and biochemical responses in plants. These stresses commonly lead to oxidative stress, cellular damage, and disruptions in metabolism, and plants activate a core set of stress-responsive genes to mitigate these disruptions, in addition to the unique and well-characterized pathways for each individual stress. Meta-analyses in various crop and model plant species have found between 20-6,000 core abiotic stress-responsive genes based on overlapping expression profiles and, in one case, support vector machine clustering. Previous efforts have focused on a narrow set of abiotic stresses, used microarrays with only partial gene representation, or had sparse gene expression datasets that fail to capture the breadth or diversity

of stress responses within a species. Here, we expanded on previous work by using diverse genotypes, abiotic stresses, tissue types, and stress severity in maize. Using both set operations and random forest classification, we found 744 core genes in all tissues and 512 in photosynthetic tissues only (Table A2.2). The number of core genes we found for drought, salt, cold, heat, low nitrogen, and flooding in maize is consistent with previous studies.

We leveraged RNA-seq data from 39 previously published experiments, downloaded from their NCBI BioProjects. Several of the previous meta-analyses analyzed the data from significantly fewer BioProjects than were used in the present study (less than 10 studies each in (Dossa et al., 2019) and (Cohen and Leach, 2019)), and to our knowledge, this is the first such meta-analysis to leverage random forest classification modeling to help identify core genes. Most previous studies have used only set operations on differentially expressed genes, which is a limited approach in that it defines core genes by their expression under stress alone. In contrast, our random forest approach identified many genes that, while differentially expressed, were not identified as core genes by the set operations method, and identified a few genes that were not differentially expressed under any condition. There was little overlap between the results from the two methods. This indicates that the random forest method identified emergent aspects of core stress response, broadening the core stress gene set. This new methodology may be informative for future core stress meta-analyses.

Gene expression changes under abiotic stress are achieved through the action of regulators such as transcription factors (TFs). We found that the core stress gene set was enriched in TFs, including particular families such as heat shock factors (HSFs), ethylene response factors (AP2/ERF-ERFs), C2C2-CO-like, NAC, and bZIP (see Table 4). These are all known to be stress-responsive TF families (Mizoi et al., 2012; Yoon et al., 2020; Andrásí et al., 2021). The HSF family are the direct regulators of heat shock proteins (HSPs) (Andrásí et al., 2021), which act as molecular chaperones under various abiotic stress conditions (ul Haq et al., 2019). Core gene members of the HSF family are present in a co-expression module with enriched GO terms related to “protein folding”, and are thus possibly involved in regulating HSPs in this module.

The members of the AP2/ERF-ERF TF family have been found to be responsive to multiple abiotic stresses, including flooding, drought, and salt (Mizoi et al., 2012; Debbarma et al., 2019). At least one member of this family, *TaERF4* from wheat, is likely a negative regulator

of salt response (Dong et al., 2012). Furthermore, gene editing of members of this family has yielded improved stress tolerance in multiple cases (Debbarma et al., 2019). The AP2/ERF-ERFs were enriched in upregulated core genes as well as upregulated heat peripheral genes for all tissues (Table A2.4), suggesting that they are important regulators of both core stress and heat-specific responses in maize. Core gene members of this family are found in a co-expression module where the only enriched GO terms are all related to ethylene (Figure A2.6C), highlighting a potential regulation by ethylene.

All of the TF families enriched and near-enriched in core genes for all tissues were present in modules that were also enriched in at least 2 stress-specific gene sets (Figure A2.6A). From this, we can conclude that, given their coexpression with both other core genes (in some modules) and peripheral genes, it is possible that these core TFs regulate both the core abiotic stress response and stress-specific responses. Thus, these core stress TFs may make good targets for improvement of abiotic stress tolerance at least in maize, and possibly in other species as well, especially considering that overexpression of stress-involved TFs has previously led to improved stress tolerance in many cases (Nakashima et al., 2012; Hoang et al., 2014; Baillo et al., 2019). However, further studies are required to elucidate the molecular mechanisms of the maize core stress genes identified in this study; for example, if a TF acts as a repressor of genes that would be beneficial if expressed under stress, overexpression of that TF could lead to a reduction in stress tolerance rather than the desired increase. This omics-only meta-analysis does not provide enough information to say definitively whether the core genes would be good targets for crop improvement, but we hope that follow-up studies will provide more information.

## REFERENCES

- Alexa A, Rahnenfuhrer J** (2022) topGO: Enrichment Analysis for Gene Ontology. R package version 2.50.0
- Andrási N, Pettkó-Szandtner A, Szabados L** (2021) Diversity of plant heat shock factors: regulation, interactions, and functions. *J Exp Bot* **72**: 1558–1575
- Aslam S, Gul N, Mir MA, Asgher M, Al-Sulami N, Abulfaraj AA, Qari S** (2021) Role of Jasmonates, Calcium, and Glutathione in Plants to Combat Abiotic Stresses Through Precise Signaling Cascade. *Front Plant Sci.* doi: 10.3389/fpls.2021.668029
- Baillo EH, Kimotho RN, Zhang Z, Xu P** (2019) Transcription Factors Associated with Abiotic and Biotic Stress Tolerance and Their Potential for Crops Improvement. *Genes* **10**: 771
- Bano C, Amist N, Singh NB** (2020) Chapter 18 - Role of polyamines in plants abiotic stress tolerance: Advances and future prospects. *In* DK Tripathi, V Pratap Singh, DK Chauhan, S Sharma, SM Prasad, NK Dubey, N Ramawat, eds, *Plant Life Chang. Environ.* Academic Press, pp 481–496
- Batista-Silva W, Heinemann B, Rugen N, Nunes-Nesi A, Araújo WL, Braun H-P, Hildebrandt TM** (2019) The role of amino acid metabolism during abiotic stress release. *Plant Cell Environ* **42**: 1630–1644
- Behdenna A, Colange M, Haziza J, Gema A, Appé G, Azencott C-A, Nordor A** (2023) pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinformatics* **24**: 459
- Calvin K, Dasgupta D, Krinner G, Mukherji A, Thorne PW, Trisos C, Romero J, Aldunce P, Barrett K, Blanco G, et al** (2023) IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. doi: 10.59327/IPCC/AR6-9789291691647
- Chen S, Zhou Y, Chen Y, Gu J** (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890
- Cohen SP, Leach JE** (2019) Abiotic and biotic stresses induce a core transcriptome response in rice. *Sci Rep* **9**: 6273
- Debbarma J, Sarki YN, Saikia B, Boruah HPD, Singha DL, Chikkaputtaiah C** (2019) Ethylene Response Factor (ERF) Family Proteins in Abiotic Stresses and CRISPR–Cas9 Genome Editing of ERFs for Multiple Abiotic Stress Tolerance in Crop Plants: A Review. *Mol Biotechnol* **61**: 153–172
- Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C** (2017)

- Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**: 316–319
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21
- Dong W, Ai X, Xu F, Quan T, Liu S, Xia G** (2012) Isolation and characterization of a bread wheat salinity responsive ERF transcription factor. *Gene* **511**: 38–45
- Dossa K, Mmadi MA, Zhou R, Zhang T, Su R, Zhang Y, Wang L, You J, Zhang X** (2019) Depicting the Core Transcriptome Modulating Multiple Abiotic Stresses Responses in Sesame (*Sesamum indicum* L.). *Int J Mol Sci* **20**: 3930
- EL Sabagh A, Islam MS, Hossain A, Iqbal MA, Mubeen M, Waleed M, Reginato M, Battaglia M, Ahmed S, Rehman A, et al** (2022) Phytohormones as Growth Regulators During Abiotic Stress Tolerance in Plants. *Front Agron*. doi: 10.3389/fagro.2022.765068
- Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S** (2020) The nf-core framework for community-curated bioinformatics pipelines. *Nat Biotechnol* **38**: 276–278
- ul Haq S, Khan A, Ali M, Khattak AM, Gai W-X, Zhang H-X, Wei A-M, Gong Z-H** (2019) Heat Shock Proteins: Dynamic Biomolecules to Counter Plant Biotic and Abiotic Stresses. *Int J Mol Sci* **20**: 5321
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, Wieser E, Taylor J, Berg S, Smith NJ, et al** (2020) Array programming with NumPy. *Nature* **585**: 357–362
- Hoang XLT, Thu NBA, Thao NP, Tran L-SP** (2014) Transcription Factors in Abiotic Stress Responses: Their Potentials in Crop Improvement. *In* P Ahmad, MR Wani, MM Azooz, L-S Phan Tran, eds, *Improv. Crops Era Clim. Chang. Vol. 2*. Springer, New York, NY, pp 337–366
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al** (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662
- Hunter JD** (2007) Matplotlib: A 2D Graphics Environment. *Comput Sci Eng* **9**: 90–95
- Langfelder P, Horvath S** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559
- Langfelder P, Horvath S** (2012) Fast R Functions for Robust Correlations and Hierarchical Clustering. *J Stat Softw* **46**: 1–17



- Li M, Sui N, Lin L, Yang Z, Zhang Y, Li M, Sui N, Lin L, Yang Z, Zhang Y** (2019) Transcriptomic profiling revealed genes involved in response to cold stress in maize. *Funct Plant Biol* **46**: 830–844
- Li P, Cao W, Fang H, Xu S, Yin S, Zhang Y, Lin D, Wang J, Chen Y, Xu C, et al** (2017) Transcriptomic Profiling of the Maize (*Zea mays* L.) Leaf Response to Abiotic Stresses at the Seedling Stage. *Front. Plant Sci.* **8**:
- Love MI, Huber W, Anders S** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550
- Ma H, Liu M** (2019) The microtubule cytoskeleton acts as a sensor for stress response signaling in plants. *Mol Biol Rep* **46**: 5603–5608
- Minocha R, Majumdar R, Minocha SC** (2014) Polyamines and abiotic stress in plants: a complex relationship1. *Front Plant Sci.* doi: 10.3389/fpls.2014.00175
- Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K** (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochim Biophys Acta BBA - Gene Regul Mech* **1819**: 86–96
- Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K** (2012) NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta BBA - Gene Regul Mech* **1819**: 97–103
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C** (2017) Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat Methods* **14**: 417–419
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al** (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**: 2825–2830
- Qi M, Liu X, Li Y, Song H, Yin Z, Zhang F, He Q, Xu Z, Zhou G** (2021) Photosynthetic resistance and resilience under drought, flooding and rewatering in maize plants. *Photosynth Res* **148**: 1–15
- R Core Team** (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, Xu J** (2019) Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. *Plants* **8**: 34
- Sanchez-Munoz R, Depaepe T, Samalova M, Hejatko J, Zaplana I, Straeten DVD** (2024) The molecular core of transcriptome responses to abiotic stress in plants: a machine

learning-driven meta-analysis. 2024.01.24.576978

- Seabold S, Perktold J** (2010) *Statsmodels: Econometric and Statistical Modeling with Python*. Austin, Texas, pp 92–96
- Shao H, Wang H, Tang X** (2015) NAC transcription factors in plant multiple abiotic stress responses: progress and prospects. *Front Plant Sci*. doi: 10.3389/fpls.2015.00902
- Sharma A, Shahzad B, Kumar V, Kohli SK, Sidhu GPS, Bali AS, Handa N, Kapoor D, Bhardwaj R, Zheng B** (2019) Phytohormones Regulate Accumulation of Osmolytes Under Abiotic Stress. *Biomolecules* **9**: 285
- Sharma M, Irfan M, Kumar A, Kumar P, Datta A** (2022) Recent Insights into Plant Circadian Clock Response Against Abiotic Stress. *J Plant Growth Regul* **41**: 3530–3543
- Shintani M, Tamura K, Bono H** (2024) Meta-analysis of public RNA sequencing data of abscisic acid-related abiotic stresses in *Arabidopsis thaliana*. *Front Plant Sci*. doi: 10.3389/fpls.2024.1343787
- Singh J, Thakur JK** (2018) Photosynthesis and Abiotic Stress in Plants. *In* S Vats, ed, *Biot. Abiotic Stress Toler. Plants*. Springer, Singapore, pp 27–46
- Soneson C, Love MI, Robinson MD** (2016) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. doi: 10.12688/f1000research.7563.2
- Tahmasebi A, Ashrafi-Dehkordi E, Shahriari AG, Mazloomi SM, Ebrahimie E** (2019) Integrative meta-analysis of transcriptomic responses to abiotic stress in cotton. *Prog Biophys Mol Biol* **146**: 112–122
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al** (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**: 261–272
- Wu Y, Li Q, Jin R, Chen W, Liu X, Kong F, Ke Y, Shi H, Yuan J** (2019) Effect of low-nitrogen stress on photosynthesis and chlorophyll fluorescence characteristics of maize cultivars with different low-nitrogen tolerances. *J Integr Agric* **18**: 1246–1256
- Yilmaz A, Nishiyama MY Jr, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E** (2009) GRASSIUS: A Platform for Comparative Regulatory Genomics across the Grasses. *Plant Physiol* **149**: 171–180
- Ying S, Webster B, Gomez-Cano L, Shivaiah K-K, Wang Q, Newton L, Grotewold E, Thompson A, Lundquist PK** (2023) Multiscale physiological responses to nitrogen supplementation of maize hybrids. *Plant Physiol* kiad583
- Yoon Y, Seo DH, Shin H, Kim HJ, Kim CM, Jang G** (2020) The Role of Stress-Responsive

Transcription Factors in Modulating Abiotic Stress Tolerance in Plants. *Agronomy* **10**: 788

**Zhang Y, Ali U, Zhang G, Yu L, Fang S, Iqbal S, Li H, Lu S, Guo L** (2019) Transcriptome analysis reveals genes commonly responding to multiple abiotic stresses in rapeseed. *Mol Breed* **39**: 158

## APPENDIX A: FIGURES AND TABLES

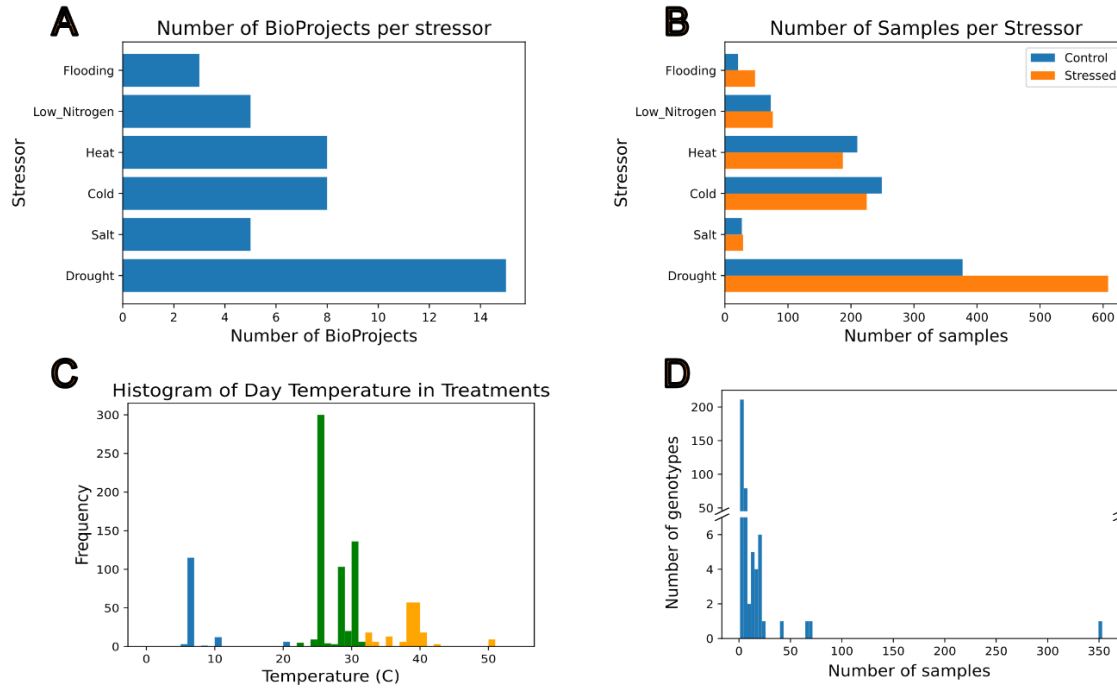


Figure A2.1: Bar plots of numbers of BioProjects (A) and samples (B) per stressor, and histograms of temperatures in different treatments (C) and number of samples in different genotypes (D). Drought has the largest number of both BioProjects and samples. Flooding has the fewest BioProjects, and salt the fewest samples. Although there is a clear distinction in the means of cold (blue), heat (orange), and control (green) treatments, there is some overlap at the edges of the ranges. Most of the genotypes have a low number of samples, with very few having more than 50.

Table A2.1: Numbers of samples in each tissue in the dataset.

Tissue	Number of samples
Leaf	1,394
Root	278
Shoot	46
Ear	44
Tassel	39
Embryo	24
Pollen	16
Stalk	10
Silk	9
Kernel	4
Ovary	4
Leaf meristem	4

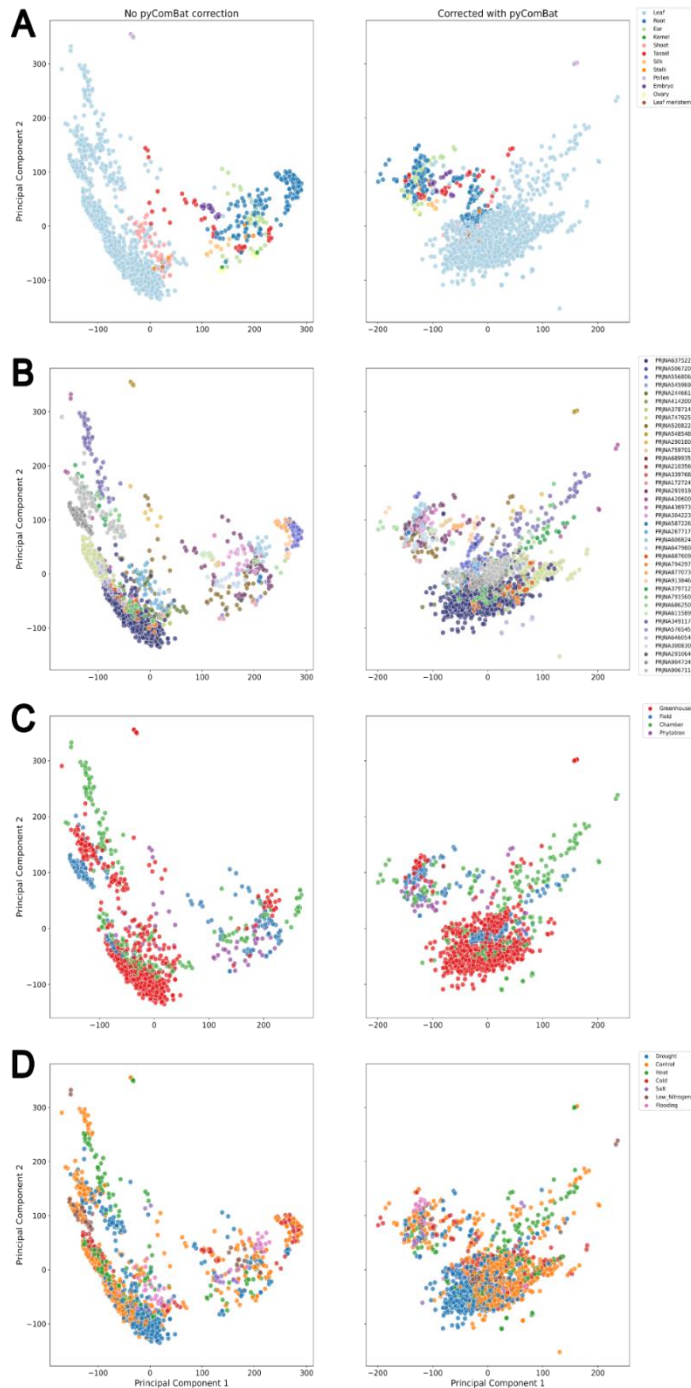
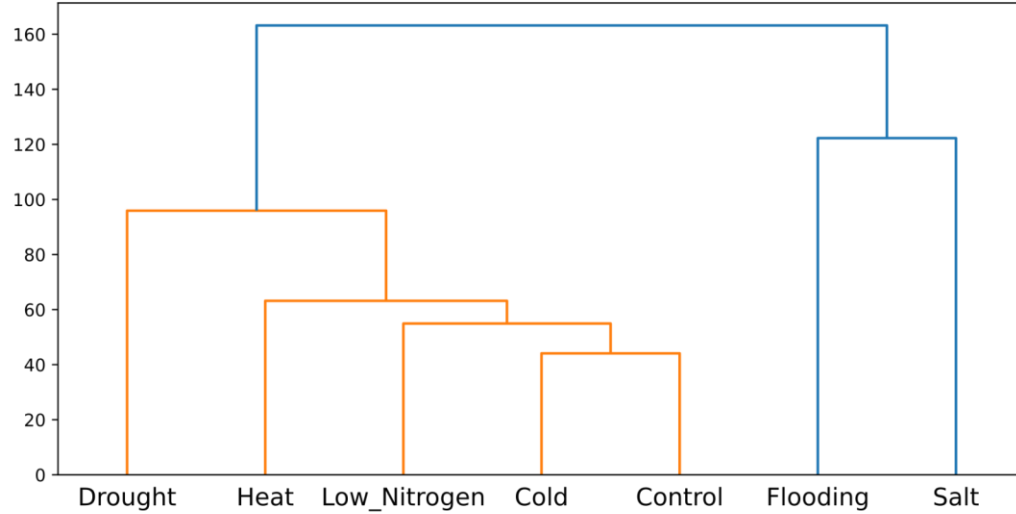


Figure A2.2: Principal component analysis biplots of log-transformed TPM, without correction for variance due to BioProject (left column) and with correction (right column). For uncorrected data, PC1 explained 28% of variance and PC2 explained 15%; for corrected data, PC1 explained 16% and PC2 14%. Biplots are colored by tissue (row A), BioProject (B), growth environment (C), and treatment (D). Samples group primarily by tissue, followed by BioProject in the uncorrected data; correction reduced but did not eliminate grouping by BioProject, possibly due to the effect of different growth environments. Core and peripheral genes were found based on corrected TPM, both for all tissues and, to reduce the effect of tissue, for photosynthetic tissues only.

**A** Hierarchical Clustering of Mean BioProject-Corrected logTPM (all samples) per Treatment



**B** Hierarchical Clustering of Mean BioProject-Corrected logTPM (photosynthetic tissues only) per Treatment

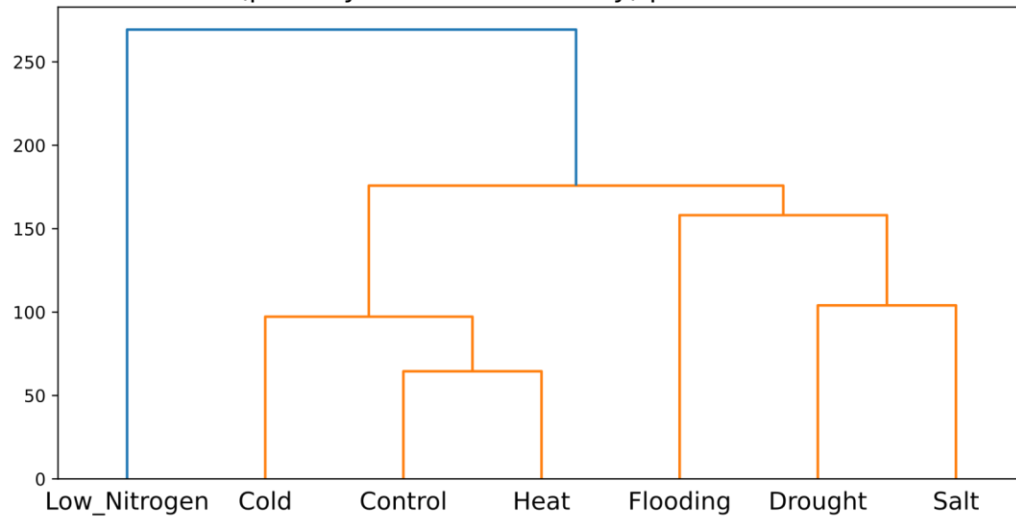


Figure A2.3: Hierarchical clustering of transcriptomes for different treatments, for all tissues (A) and photosynthetic tissues only (B).

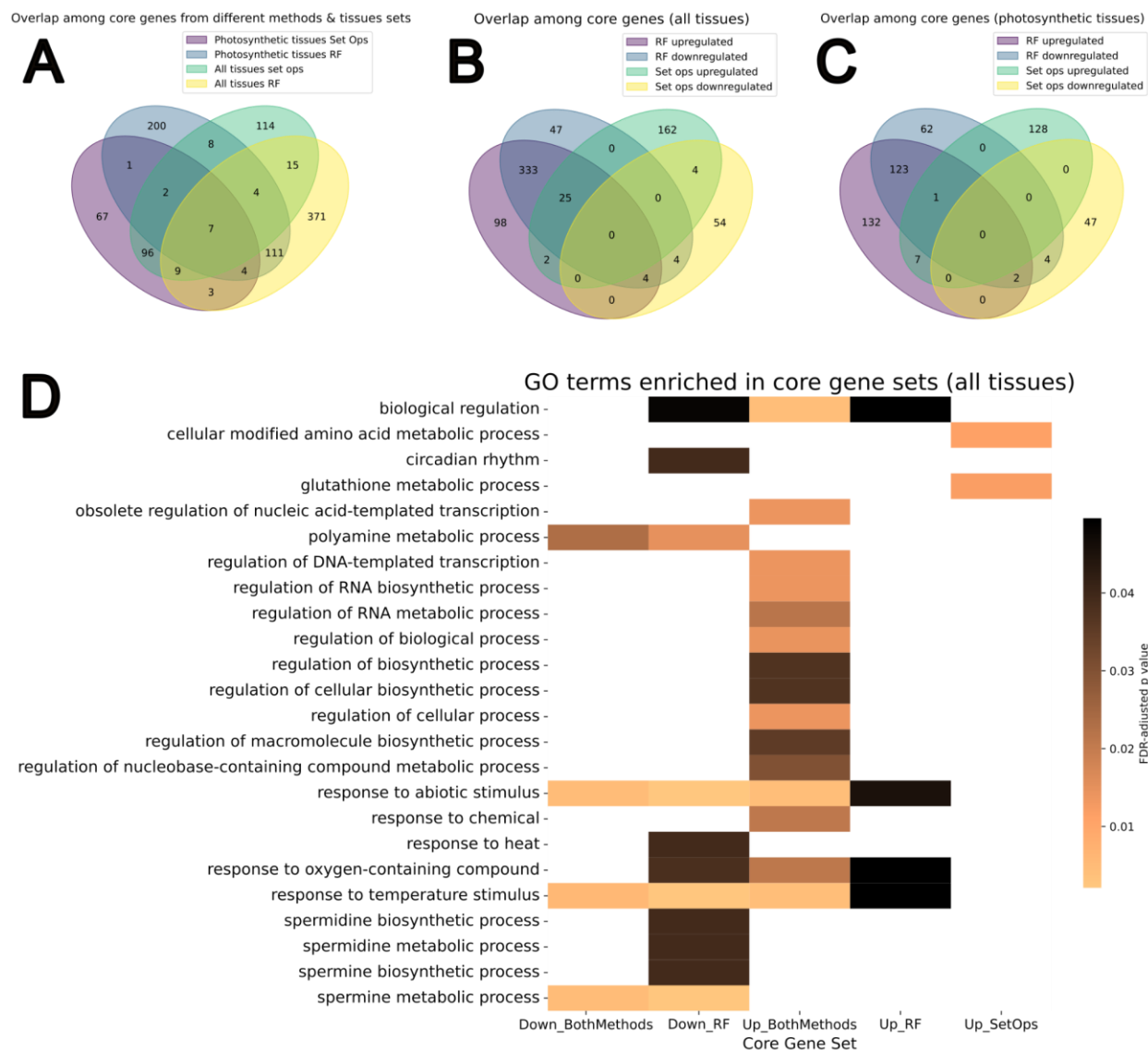


Figure A2.4: Basic characterization of the core gene sets. Set operations and random forest identified largely distinct core gene sets, and core genes from different tissue sets also had little overlap (A). For both all tissues (B) and photosynthetic tissues (C), there was more overlap between upregulated and downregulated core genes (in any stress condition) for genes identified by random forest than for those identified by set operations (“set ops”). A number of Gene Ontology (GO) terms related to regulatory processes, metabolites, and stress response were enriched in core gene sets from all tissues (D); no enriched GO terms were found for core genes from photosynthetic tissues.



Table A2.2: Summary of core gene numbers by method, tissue set, and regulation. Overall, more core genes were upregulated than downregulated, more core genes were found by random forest than by set operations, and more core genes were found in all tissues than in photosynthetic tissues only. Note that some of the core genes identified by random forest are upregulated in some stress conditions and downregulated in others.

Set of tissues	Method	Number of upregulated genes	Number of downregulated genes	Number of genes with no differential expression	Total genes
All tissues	Set operations	193	66	0	255
	Random forest	462	413	11	524
	Combined	628	471	11	744
Photosynthetic tissues	Set operations	136	53	0	189
	Random forest	265	192	6	337
	Combined	393	239	6	512

Table A2.3: Numbers of peripheral genes, i.e. genes only differentially expressed under a single stress condition, in different tissue sets. Genes differentially expressed in at least one study of the given stressor were included, making this a broad survey. Flooding consistently has the fewest peripheral genes, probably because it also has the fewest studies (see Figure 1A).

Set of tissues	Stressor	Number of upregulated genes	Number of downregulated genes	Total genes
All tissues	Cold	3,709	3,691	7,065
	Low nitrogen	1,355	1,487	2,825
	Heat	2,780	485	3,238
	Drought	1,599	3,609	5,092
	Flooding	390	254	643
	Salt	753	422	1,170
Photosynthetic tissues	Cold	817	2,394	3,207
	Low nitrogen	1,740	1,260	2,992
	Heat	2,849	1,265	4,079
	Drought	2,352	2,780	4,922
	Flooding	394	162	556
	Salt	600	413	1,009

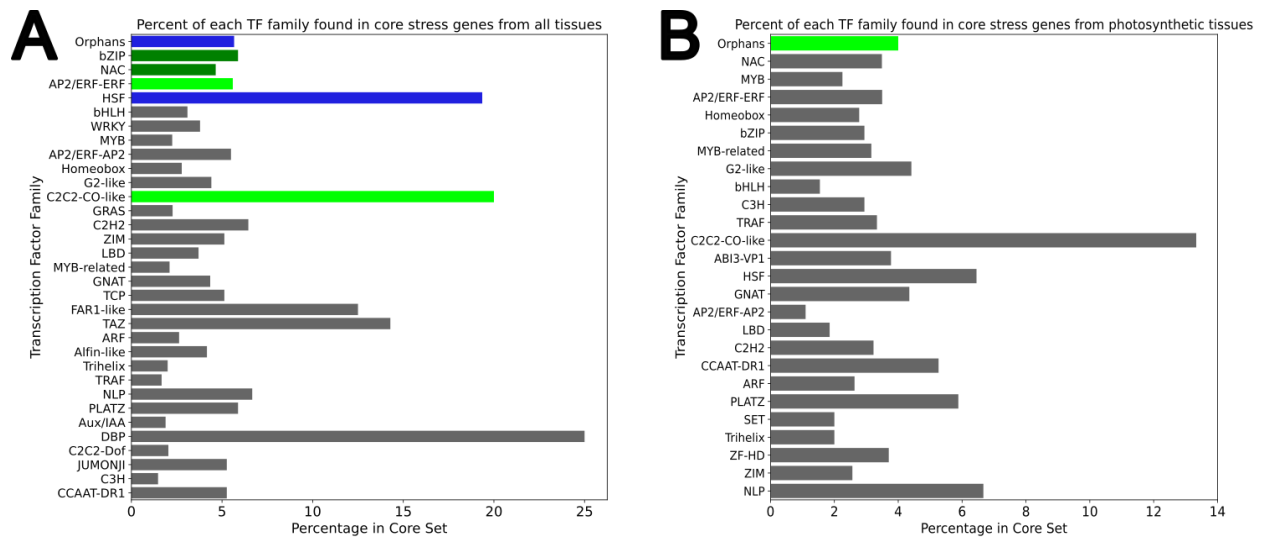


Figure A2.5: Percentages of transcription factor families found in the core gene sets from all tissues (A) and photosynthetic tissues (B). Light green indicates that a given TF family is enriched ( $p < 0.05$ , Fisher's exact test) in upregulated core genes, dark green that it is near-enriched ( $0.05 < p < 0.1$ ) in upregulated genes, and blue that it is enriched in both upregulated and downregulated core genes. Families not enriched or near-enriched in either core gene set are depicted in gray.

Table A2.4: Summary of TF families enriched (adjusted  $P < 0.05$ ) and near-enriched ( $0.05 < P < 0.1$ ) in core gene sets. Notably, non-orphan TF families were found to be enriched and near-enriched only in the set of core genes from all tissues, not in photosynthetic tissues only. Enriched and near-enriched non-orphan families in all tissues are mostly known to be related to abiotic stress response.

Set of tissues	Regulatory direction	TF family	P-value
All tissues	Upregulated	AP2/ERF-ERF	0.0468
		C2C2-CO-like	0.0375
		HSF	0.005
		Orphans	0.005
		NAC	0.0924
		bZIP	0.0924
	Downregulated	HSF	0.0191
		Orphans	0.007
Photosynthetic tissues	Upregulated	Orphans	0.0847

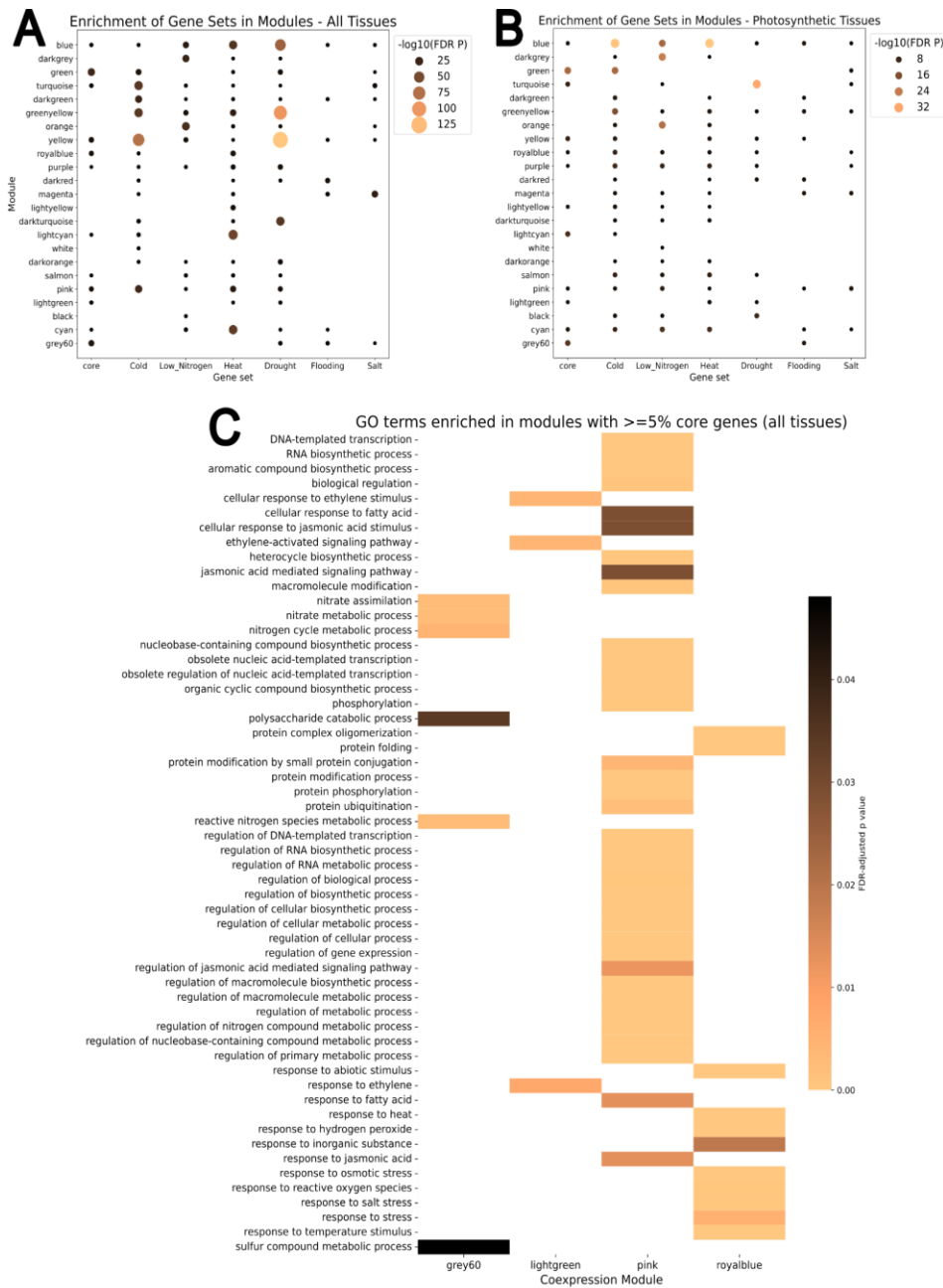


Figure A2.6: Bubble plots of gene set enrichment in different coexpression modules for core and peripheral gene sets from all tissues (A) and photosynthetic tissues (B). The negative log<sub>10</sub> of the FDR-adjusted p-value is plotted, so larger bubbles indicate more highly significant enrichment. Non-significant enrichments were not plotted. None of the modules enriched in core genes were also not enriched in at least 1 set of peripheral genes; however, there are several modules enriched in all six peripheral gene sets.

APPENDIX B: SUPPLEMENTAL FIGURES AND TABLES

Table B2.1: Salmon mapping rates of two genotypes of maize when mapped to their own genomes and to B73. Mapping rates in B73 were comparable to own genome mapping rates.

Genotype	Mean mapping rate against self	Mean mapping rate against B73
Oh43	89.08%	85.45%
CML69	83.86%	79.62%

### Histogram of mapping rates for SRA samples

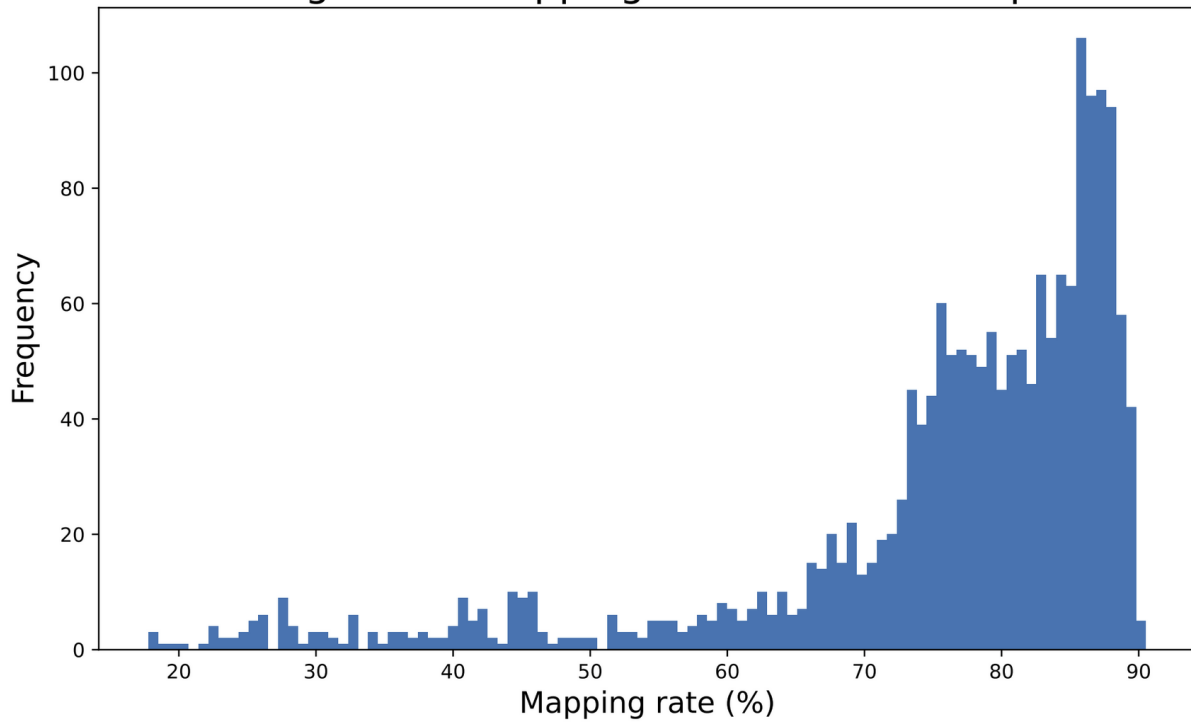


Figure B2.1: Salmon pseudoalignment mapping rates for all samples downloaded from the NCBI Sequence Read Archive and processed against the B73 V5 genome. Most samples had mapping rates of greater than 60%.

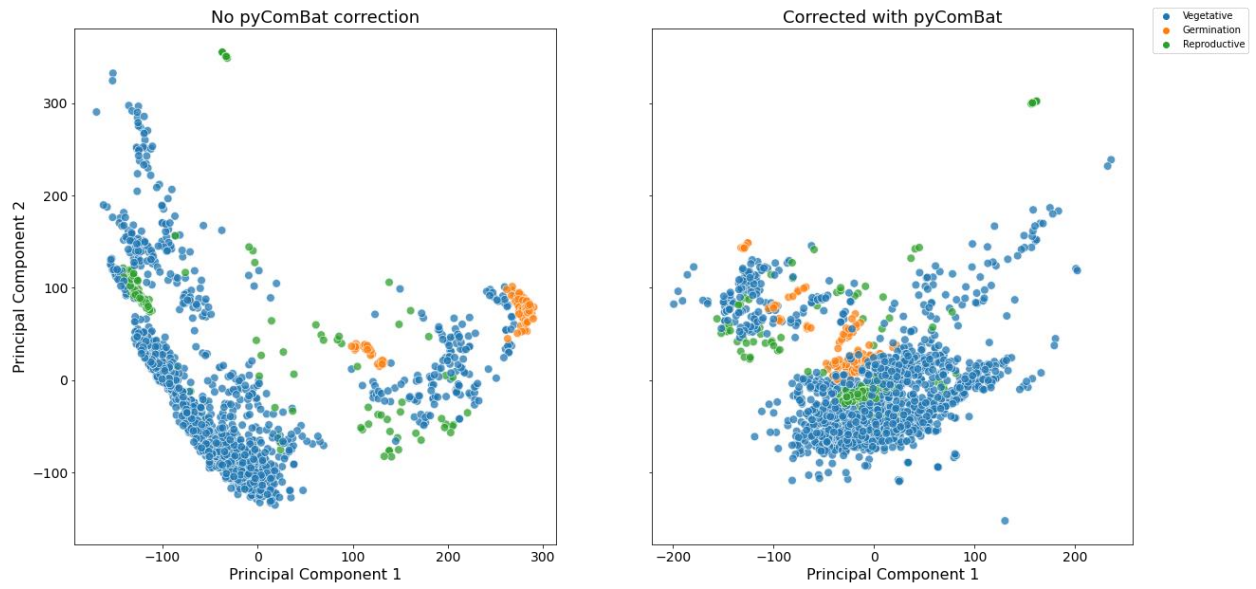


Figure B2.2: Principal component analysis biplot colored by developmental stage (vegetative, reproductive, or germination). Uncorrected  $\log(\text{TPM})$  are on the left, corrected on the right.



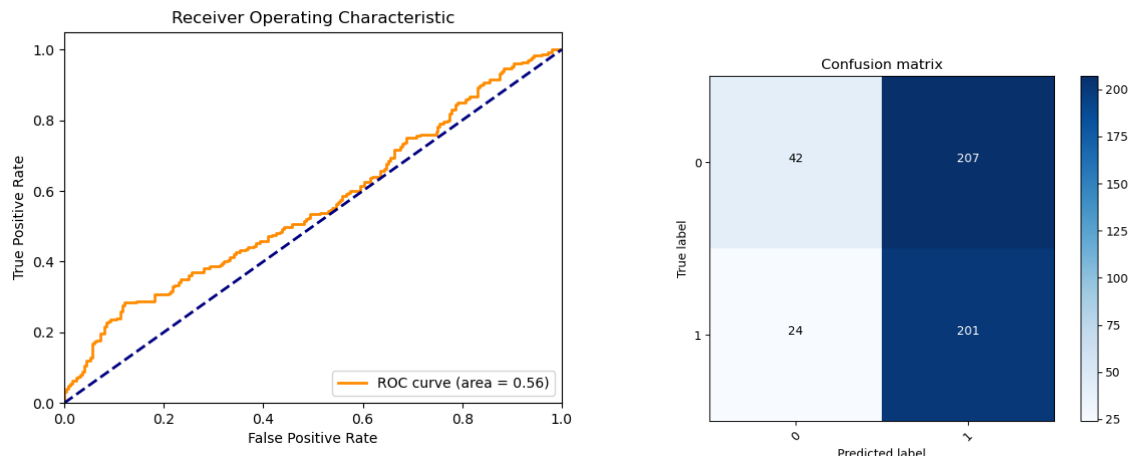


Figure B2.3A: ROC curve and confusion matrix for the model with all tissues cold as the test data.

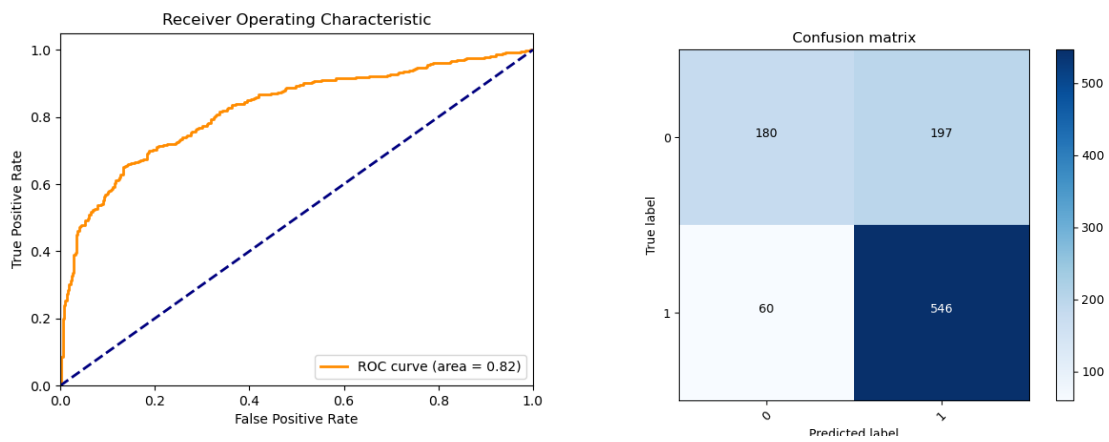


Figure B2.3B: ROC curve and confusion matrix for the model with all tissues drought as the test data.

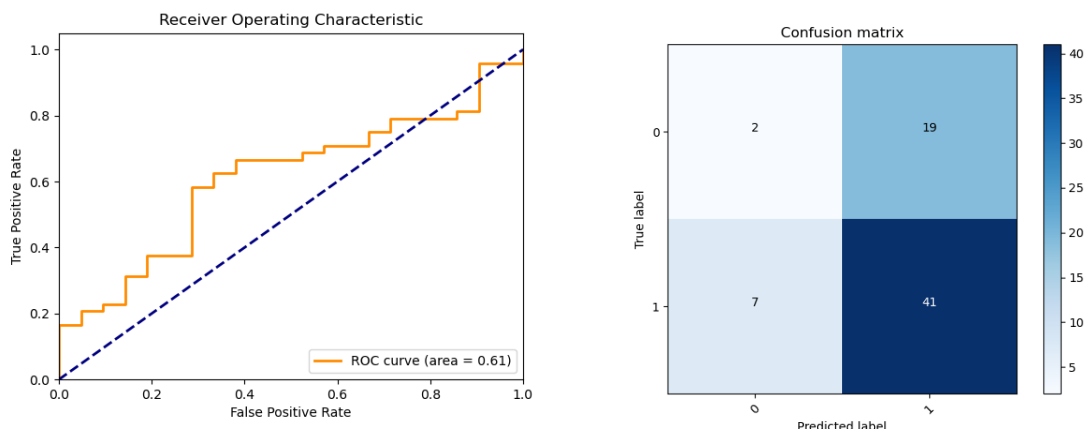


Figure B2.3C: ROC curve and confusion matrix for the model with all tissues flooding as the test data.

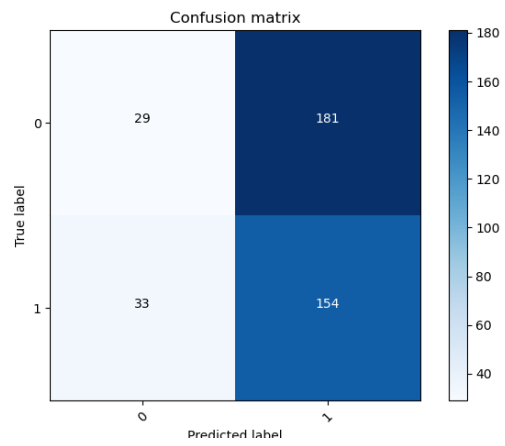
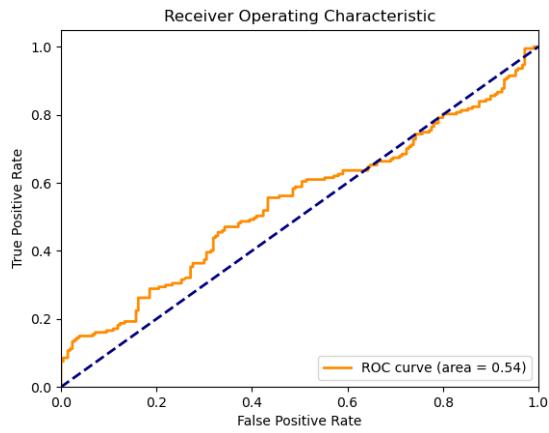


Figure B2.3D: ROC curve and confusion matrix for the model with all tissues heat as the test data.

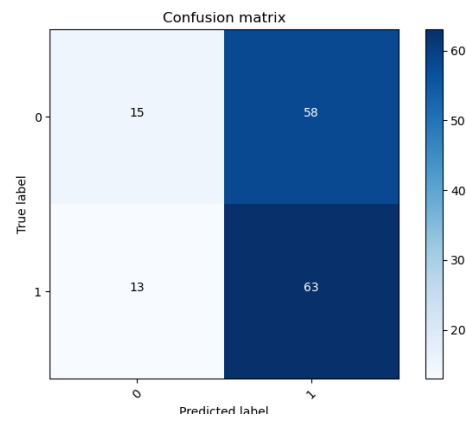
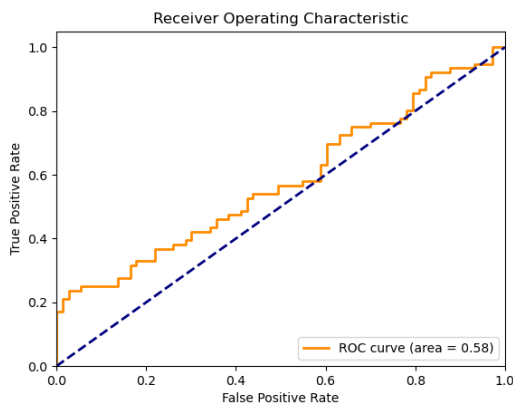


Figure B2.3E: ROC curve and confusion matrix for the model with all tissues low nitrogen as the test data.

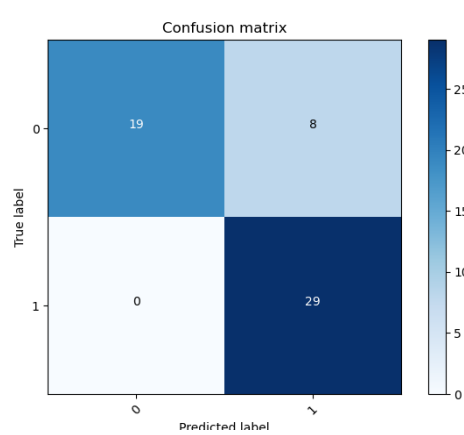
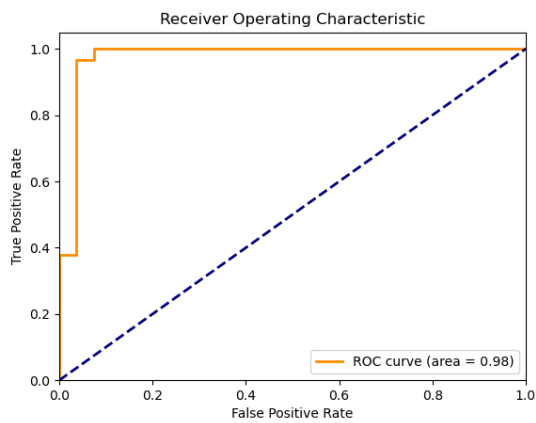


Figure B2.3F: ROC curve and confusion matrix for the model with all tissues salt as the test data.

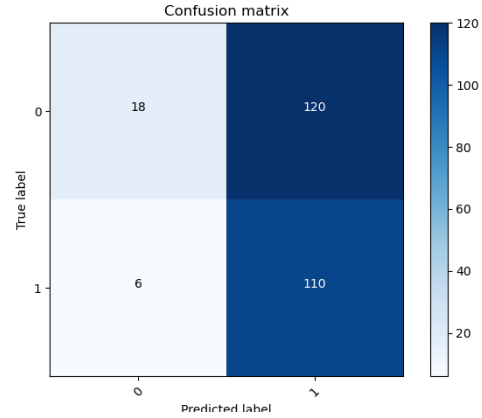
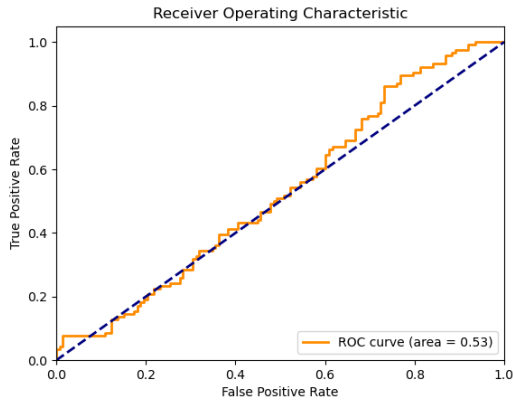


Figure B2.4A: ROC curve and confusion matrix for model with photosynthetic tissues cold as test data.

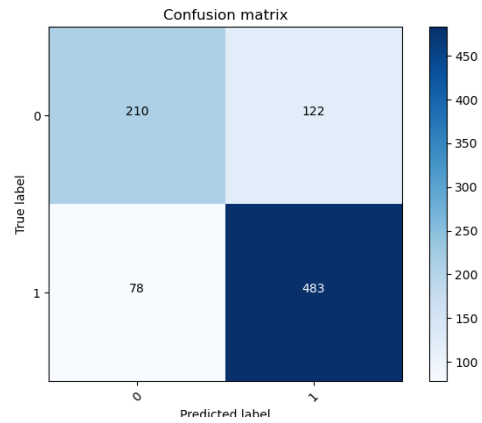
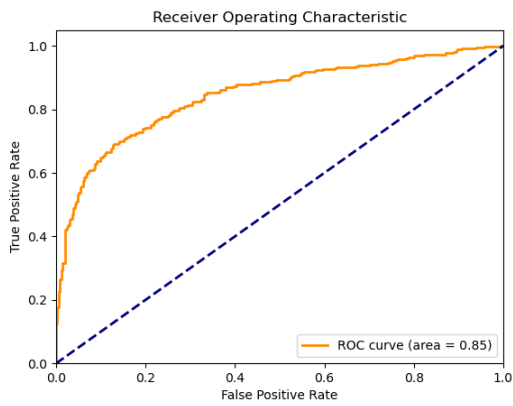


Figure B2.4B: ROC curve and confusion matrix for model with photosynthetic tissues drought as test data.

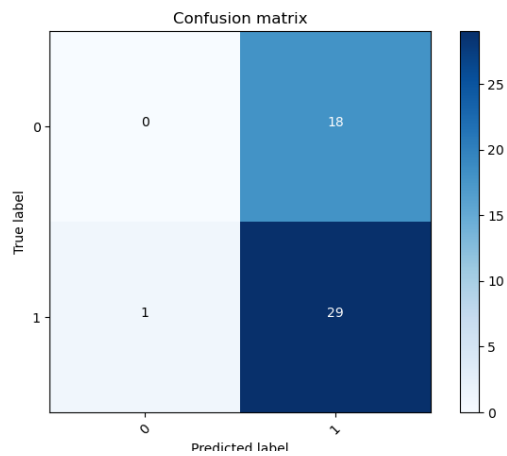
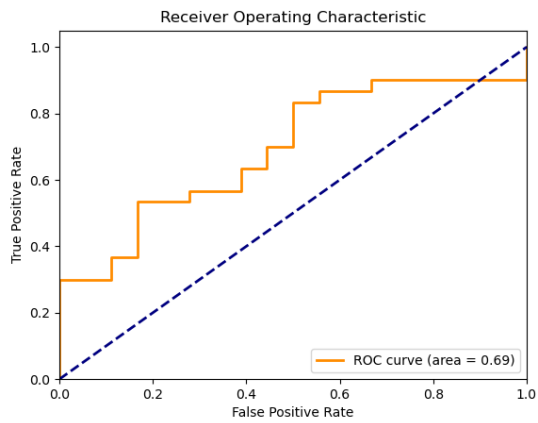


Figure B2.4C: ROC curve and confusion matrix for model with photosynthetic tissues flooding as test data.

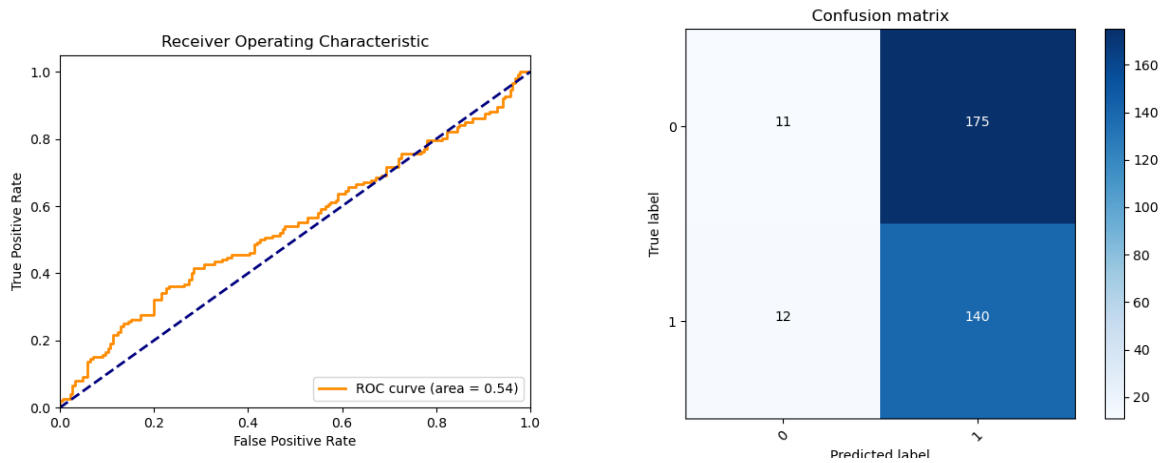


Figure B2.4D: ROC curve and confusion matrix for model with photosynthetic tissues heat as the test data.

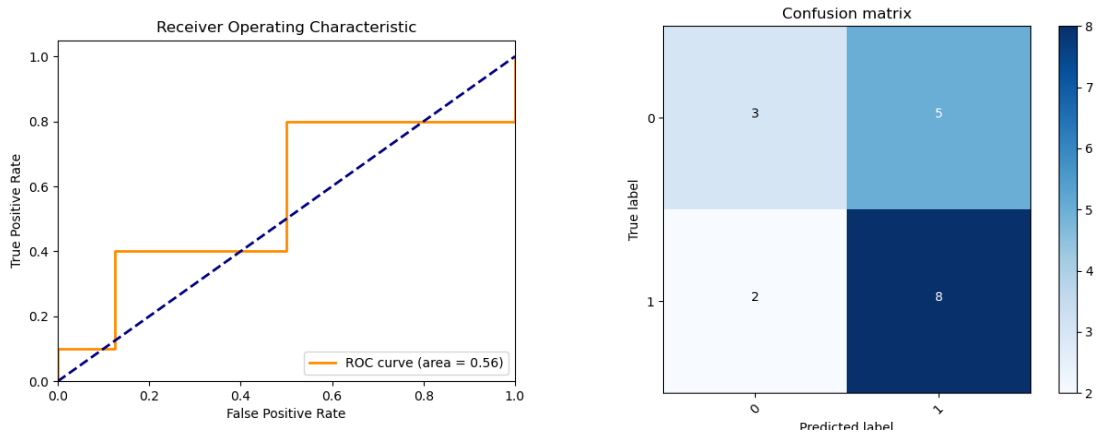


Figure B2.4E: ROC curve and confusion matrix for model with photosynthetic tissues low nitrogen as the test data.

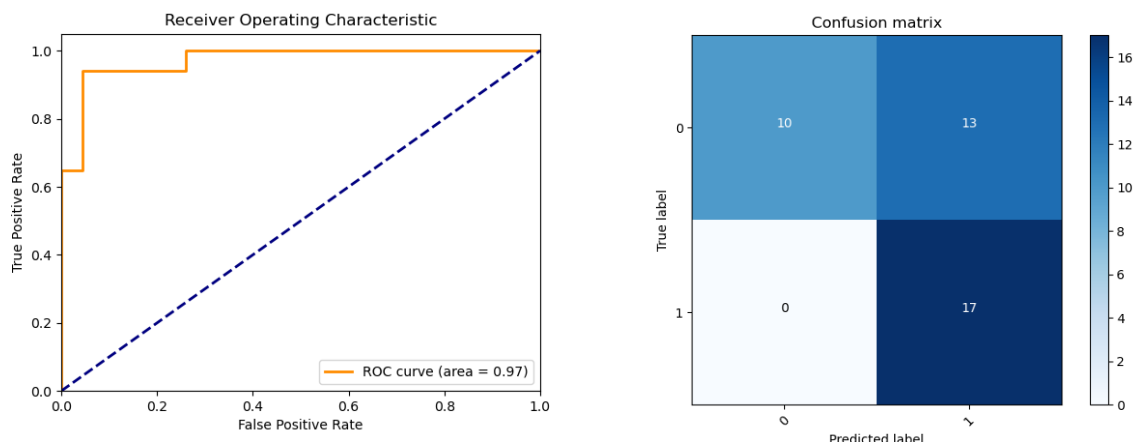


Figure B2.4F: ROC curve and confusion matrix for model with photosynthetic tissues salt as the test data.

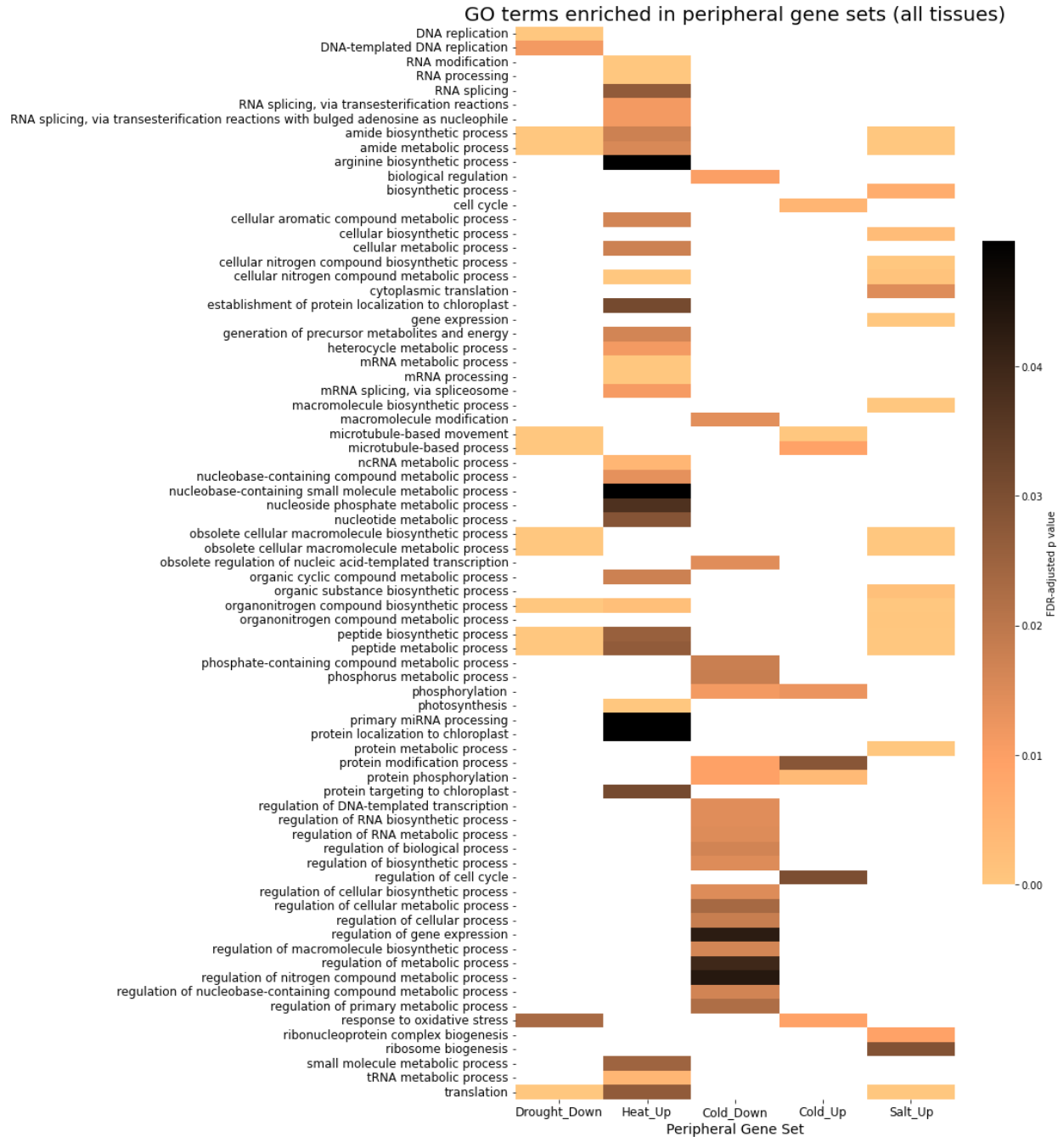


Figure B2.5: Enriched GO terms in peripheral gene sets for all tissues. Lighter colors indicate higher enrichment. Up, upregulated; Down, downregulated. Only the five peripheral gene sets shown had any GO term enrichment.

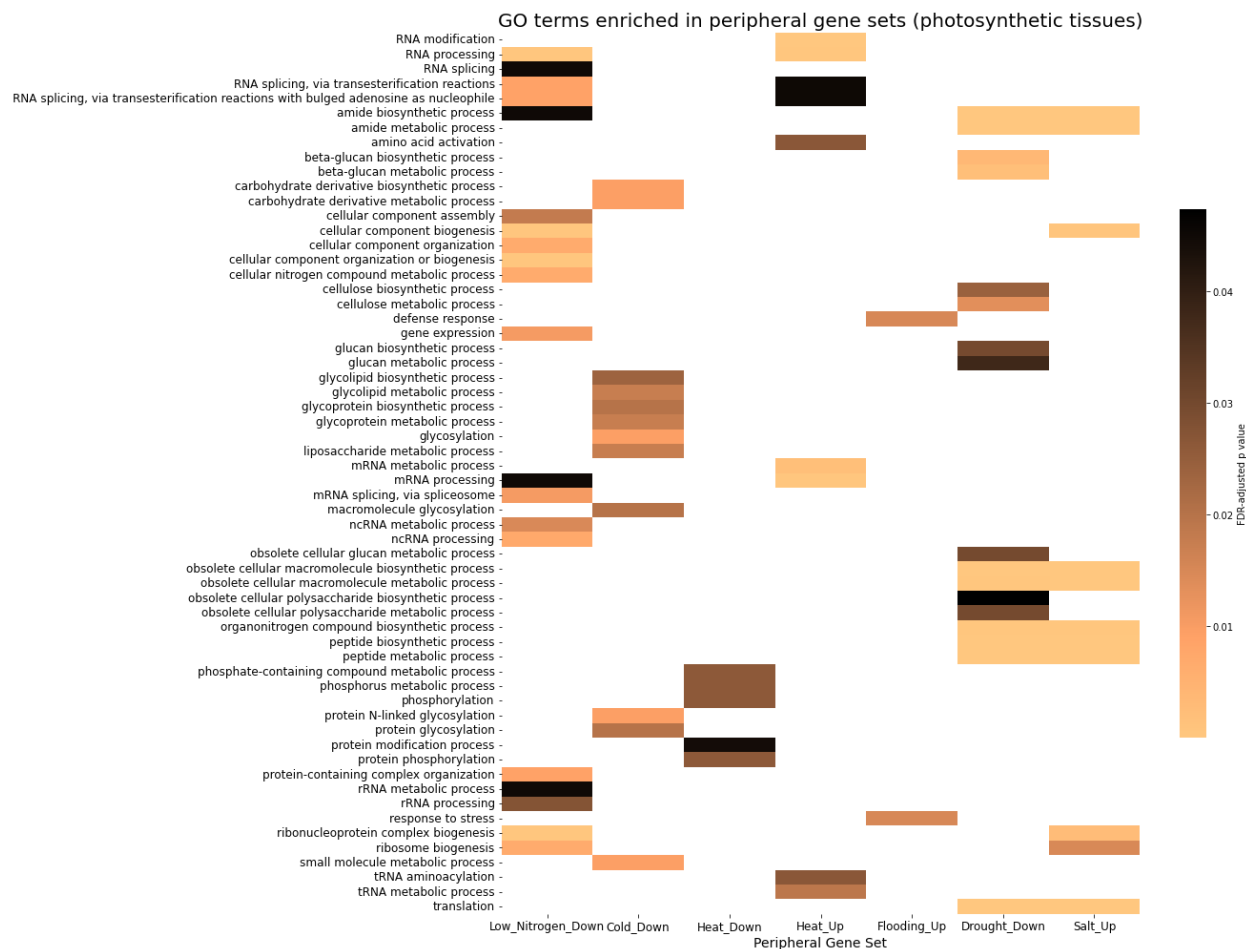


Figure B2.6: Enriched GO terms in peripheral genes from photosynthetic tissues. A lighter color indicates higher enrichment (lower P-value), except white which indicates no enrichment. Only the gene sets listed on the X axis had any enriched GO terms.

Table B2.2: TF families enriched in different peripheral stress gene sets for both sets of tissues. P-values given are from Fisher's exact test, corrected with FDR.

Set of tissues	Stressor	Regulatory direction	TF family	P-value
All tissues	Cold	Upregulated	HMG	0.0424
			OVATE	0.0424
			SHI/STY (SRS)	0.0094
		Downregulated	C2H2	0.0749
			WRKY	0.0749
	Low nitrogen	Upregulated	CCAAT-HAP2	0.0619
			ZF-HD	0.0234
	Heat	Upregulated	SWI/SNF-SWI3	0.0022
			mTERF	0.0775
		Downregulated	ARF	<0.0001
			CCAAT-HAP2	0.0012
			SBP	0.002
			Sigma70-like	0.0661
	Drought	Upregulated	C2C2-YABBY	0.0909
		Downregulated	CCAAT-DR1	0.0623
	Flooding	Upregulated	C2C2-CO-like	0.0283
		Downregulated	MYB	0.0087
NAC			0.0385	
Photosynthetic	Cold	Downregulated	GNAT	0.0168

Table B2.2 (cont'd)

			SET	0.0156
	Low nitrogen	Upregulated	CCAAT-HAP2	0.0023
	Heat	Upregulated	SET	0.017
			SWI/SNF-SWI3	0.0626
		Downregulated	ARF	0.032
			FAR1-like	0.032
			NAC	0.0804
			SBP	0.0011



### **Chapter 3: Improved genome assembly and annotation for *Eragrostis nindensis*, an autotetraploid desiccation tolerant grass**

#### **ABSTRACT**

Complete genome assemblies and annotations enable research on plant traits including desiccation tolerance, yet wild plants are under-sequenced compared to crops, and few wild species have improved genome versions beyond the first published. Here, we present the improved genome assembly and annotation of *Eragrostis nindensis*, a wild desiccation tolerant grass in the same genus as the staple Ethiopian cereal crop teff. The assembly presented here (V3) is 20 times more contiguous than the previous version of the assembly (V2.1), and has better representation of the subgenomes. The updated annotation better captures the gene space with 78,612 gene models and improved alignment of gene expression data. This improved genome enabled us to study the polyploid origin of *E. nindensis*, and we determined that it is likely an autotetraploid. These genomic resources will facilitate research on the genomics of desiccation tolerance in this important species, with potential applications to teff and other cereals.

#### **INTRODUCTION**

In the current era, availability of quality genomic resources for plant species is more important than ever before. Wild species, in particular, have been severely under-sequenced compared to crops and model organisms (Marks et al., 2021). Polyploidy, which is highly prevalent in the resilient grass subfamily Chloridoideae (Roodt and Spies, 2003), makes quality genome assembly more challenging. Polyploid genomes may be allopolyploid, with two or more sets of homeologs resulting from a hybridization event followed by whole-genome duplication, or autopolyploid, with sets of homeologs resulting from a spontaneous whole-genome duplication. Autopolyploid genomes are particularly difficult to assemble to chromosome scale, although this has been achieved for sugarcane (Zhang et al., 2018) and rhubarb (Zhang et al., 2024). Long read sequencing using technologies such as PacBio has helped enable assembly of complex plant genomes (Michael and VanBuren, 2020).

Vegetative desiccation tolerance is the ability of certain plant species to equilibrate their water content with the humidity of the surrounding air (Oliver et al., 2000). These species, commonly referred to as “resurrection plants,” typically live on rocky outcrops with marginal soil conditions and/or in areas with seasonal dryness and can remain desiccated for extended

periods of time before fully rehydrating (Porembski and Barthlott, 2000). *Eragrostis nindensis* is a desiccation tolerant grass native to southern Africa. Although desiccation tolerance is present across many angiosperm families, *E. nindensis* is the only well-studied resurrection plant in the same genus as a cereal crop, *E. tef* (Pardo et al., 2020), which is a staple in Ethiopia and Eritrea. Thus, discoveries made about desiccation tolerance in *E. nindensis* may be particularly applicable to improving the resilience of *E. tef* and other cereals.

A draft genome assembly for *E. nindensis* was published in 2020 (Pardo et al., 2020) but complex polyploidy and technological limitations resulted in a highly fragmented assembly with partially assembled haplotypes. In this work, we re-sequenced the genome of *E. nindensis* with improved PacBio HiFi technology, significantly improving its contiguity, and improved the annotation as well. We also investigated the polyploid origin of *E. nindensis*, which is known to be a tetraploid (Pardo et al., 2020), and suggest that it is an autopolyploid.

## METHODS

### *Plant materials and growth conditions*

High molecular weight genomic DNA was collected from a single *Eragrostis nindensis* plant that was originally obtained from the USDA Germplasm Resources Information Network ([www.ars-grin.gov](http://www.ars-grin.gov); accession PI 410063). The plant was maintained at 28/22C day/night temperatures and under a 12 hour photoperiod. Leaf tissue was collected and immediately frozen in liquid nitrogen.

For RNA collection, *E. nindensis* plants were grown from seeds under identical conditions as described above. Three tissue types were collected: leaves and roots from seedlings, and roots from more mature plants. At collection, roots were quickly rinsed in water before being frozen in liquid nitrogen. For seedlings, the entire root system was collected; for adult roots, several root tip sections were collected. Leaves were immediately frozen in liquid nitrogen.

### *Nucleic acid extraction, library preparation, and sequencing*

High molecular weight genomic DNA of *E. nindensis* was extracted from leaf tissue using a modified nuclei prep (Zhang et al., 1995). PacBio high fidelity (HiFi) library preparation and sequencing was performed at the University of Georgia Genomics and Bioinformatics Core. DNA quality was verified using a Qubit, Nanodrop, and Fragment Analyzer, and the SMRTbell Express template prep kit v2.0 (Pacific Biosciences, Menlo Park, CA) was used for circular

consensus sequencing library construction. The library was size selected using BluePippin (Sage Science, Beverly, MA) with a target a library size of >10 kb. Sequencing was performed on the PacBio Sequel II platform and ~60 Gb of circular consensus sequencing (CCS) reads were generated.

Total RNA was extracted from ground seedling root and leaf tissue along with adult root tissue using a Trizol method with the Zymo Direct-zol RNA Miniprep Plus kit according to the manufacturer's instructions. The RNA was quantified with the Qubit based RNA Broad Range before sequencing. Full length c-DNA library preparation and sequencing were performed at the MSU RTSF Genomics facility in East Lansing, MI. Briefly, a cDNA-PCR library was generated for each sample, and these were sequenced using an Oxford Nanopore GridION instrument.

#### *Genome assembly*

*E. nindensis* is a complex tetraploid, and the genome size and within-genome heterozygosity were estimated using a k-mer based approach. K-mers were counted using the HiFi reads with Jellyfish v2.3.0 (Marçais and Kingsford, 2011) and the k-mer count distribution was used to model the heterozygosity, genome size, and putative polyploid origin using GenomeScope v2.0 (Vurture et al., 2017; Ranallo-Benavidez et al., 2020). HiFi reads were corrected and assembled using hifiasm v0.18 (Cheng et al., 2021; Cheng et al., 2022), which is optimized for haplotype resolved assembly of highly accurate long reads. 59.5 Gb of raw HiFi reads collectively representing 79x coverage of the haploid *E. nindensis* genome were assembled by hifiasm with the following modified parameters : '-l 2 -s 0.1'. The haplotype 1 assembly consisting of 828 contigs with a total length of 879 Mb and contig N50 of 10.4 Mb was used for downstream analyses.

#### *Full length cDNA data processing*

The Nanopore RNA-seq data were first filtered using NanoPack (De Coster et al., 2018) to remove reads less than 150 bp in length. Next, the remaining reads were trimmed with Porechop v0.2.4 (<https://github.com/rrwick/Porechop>) and mapped to the genome assembly using minimap2 v2.18 (Li, 2018) and SAMtools v1.11 (Danecek et al., 2021). The *de novo* transcriptome was built using StringTie2 v2.1.3 (Kovaka et al., 2019).

#### *Genome annotation*

The transposable elements in the genome assembly were annotated using EDTA (Ou et al., 2019). This was followed by three rounds of annotation with MAKER-P v2.31.10 (Campbell

et al., 2014); the first round did not include gene prediction. For the second and third rounds of annotation, gene prediction was performed with SNAP (Korf, 2004) and AUGUSTUS v3.3.2 (Stanke et al., 2008). For all three rounds, the Nanopore RNA-seq data described above was used as transcript evidence. Protein evidence was derived from the following species: *Arabidopsis thaliana* Araport11 (Cheng et al., 2017), *Oryza sativa* v7.0 (Ouyang et al., 2007), *Zea mays* v5 (Hufford et al., 2021), *Oropetium thomaeum* v2.1 (VanBuren et al., 2018), *E. tef* v3.1 (VanBuren et al., 2020), *Sorghum bicolor* v3.1.1 (McCormick et al., 2018), *Setaria italica* v2.2 (Bennetzen et al., 2012), and *Brachypodium distachyon* v3.2 (Vogel et al., 2010). Transposase genes in *E. nindensis* were identified by BLAST to a previously defined database of transposases, with an E value cutoff of 1e-20. The identified transposases were filtered out of the annotation.

The annotation was benchmarked using the Embryophyta database of Basic Universal Single Copy Orthologs (BUSCO; Manni et al., 2021) and with the LTR Assembly Index (LAI), an index of genome assembly quality based on retrotransposon content which was found using the program LTR\_retriever (Ou et al., 2018; Ou and Jiang, 2018).

#### *Functional annotation*

InterProScan v5.57-90.0 (Jones et al., 2014) was run on the *E. nindensis* V3 genome, using the Pfam (Mistry et al., 2021), TIGRFAM (Li et al., 2021), and Gene Ontology (GO; Ashburner et al., 2000; The Gene Ontology Consortium et al., 2021) databases. This output was combined with the Pfam domain output of hmmsearch, which was run using HMMER v3.3.2 (Eddy, 2011). Further GO terms were identified by reciprocal BLAST to *A. thaliana* and extraction of GO terms from the TAIR database (Berardini et al., 2004).

#### *Mapping of pre-existing RNA-seq data*

Illumina RNA-seq data were obtained from an unpublished desiccation experiment. Twelve samples were trimmed with fastp (Chen et al., 2018). Following trimming, samples were pseudoaligned to the V2.1 and V3 reference transcriptomes using Kallisto (Bray et al., 2016) and aligned to the V2.1 and V3 reference genomes using HISAT2 (Kim et al., 2019). Mapping rates were extracted from the output of each program and mean mapping rate for all twelve samples was compared between the two genome versions as a quality metric.

#### *Synteny*

MCSan Python (Tang et al., 2008) was used to find the syntenic orthologs between *E. nindensis* V3 and *E. tef* V3 (VanBuren et al., 2020), as well as between *E. nindensis* V2.1 (Pardo

et al., 2020) and *E. tef* V3 (for comparison to the *E. nindensis* V3 synteny results). In addition, 1:1 syntenic orthologs were found between *E. nindensis* V2.1 and V3, to aid in future conversion between V2.1 and V3 gene IDs.

### *Polyploid origin*

To determine the polyploid origin of *E. nindensis*, we plotted the *k*-mer spectrum of the assembly using Jellyfish (Marçais and Kingsford, 2011) and GenomeScope (Vurture et al., 2017; Ranallo-Benavidez et al., 2020) and compared it with the spectrum patterns in (Becher et al., 2020). We also calculated *K<sub>s</sub>* for the *E. nindensis* V3 and *E. tef* genomes, plotted their histograms, and compared them (*E. tef* is a known allotetraploid; VanBuren et al., 2020). *K<sub>s</sub>* was calculated and mixed modeling conducted to identify *K<sub>s</sub>* peaks using wgd v2 (Chen and Zwaenepoel, 2023).

## RESULTS

### *Genome assembly and annotation*

*E. nindensis* is a complex polyploid, and the previously published genome assembly (V2.1) was sequenced using early PacBio sequencing, resulting in a relatively low contiguity (contig N50 of 0.52 Mb) and poor subgenome resolution (Pardo et al., 2020). We sought to improve the contiguity and haplotype resolution of the tetraploid *E. nindensis* genome using PacBio high-fidelity (HiFi) technology which uses a circular consensus sequencing approach to improve read accuracy to ~99.9%. These long and highly accurate reads can help resolve complex heterozygous and polyploid plant genomes with haplotype resolution (Michael and VanBuren, 2020). Approximately 79x genome coverage (59.5 Gb) of HiFi reads were assembled using hifiasm (Cheng et al., 2021), which resulted in a greatly improved contig N50 of 10.43 Mb in V3 compared to an N50 of 0.52 Mb in V2.1. The final V3 assembly contains a mixture of 2-3 haplotypes yielding a reduced total assembly size of 897 Mb compared to the V2.1 assembly size of 986 Mb (Table A3.1), both of which are comparable to the estimated genome size of 1.0 Gb (Pardo et al., 2020). There are many possible reasons for this decrease in assembly size; it is most likely that the low error rate of HiFi sequencing collapsed some haplotypes in V3 that were uncollapsed in the previous version of the genome.

LTR Assembly Index (LAI; Ou et al., 2018), a measure of assembly continuity utilizing LTR retrotransposons, was also improved in V3 over V2.1, with a whole genome LAI of 23.64 in V3 compared to 21.79 in V2.1. Furthermore, we analyzed the synteny of *E. nindensis* V2.1

and V3 compared to *E. tef*; if we had assembled only 2 haplotypes in *E. nindensis*, we would expect a 2:2 pattern of synteny between *E. nindensis* and *E. tef*, since both species are tetraploid. In the V2.1 assembly, only 37% of gene models had a syntenic depth pattern of 2:2 against *E. tef*, compared to 64% in V3, a clear improvement (Figure A3.1). In V3, furthermore, 11% of gene models had a 1:2 pattern and 16% had a 3:2 pattern, whereas in V2.1 18% had a 1:2 and 24% had a 3:2 pattern.

*E. nindensis* V3 was annotated with the MAKER-P pipeline (Campbell et al., 2014), including AUGUSTUS and SNAP for gene prediction (Korf, 2004; Stanke et al., 2008). In V2.1, over 100,000 gene models were annotated; in V3, 78,612 gene models were predicted. This reduction in number of gene models may be due to excessive fragmentation of gene models in V2.1; we found that 87,017 V2.1 genes had 1:1 syntenic orthology to 43,993 V3 genes, indicating significant collapsing of previously fragmented gene models. The percentage of complete BUSCO was slightly improved in V3 (92.8% vs. 92.1%; Table 1). In addition, mean RNA-seq mapping rates using Kallisto (Bray et al., 2016) were improved for V3 (71.5%) compared to V2.1 (69.9%).

#### *Polyploid origin of E. nindensis*

*E. nindensis*, like most chloridoid grasses, is polyploid ( $2n = 4x = 40$ ) (Roodt and Spies, 2003), but its polyploid origin was unclear based on the V2.1 genome (Pardo et al. 2020). We used a combination of *k*-mer distributions and divergence between homeologs to test if *E. nindensis* is an auto- or allotetraploid. We calculated the number of synonymous substitutions per synonymous site (*K*<sub>s</sub>) between homeologous gene pairs across the *E. nindensis* V3 genome and compared the *K*<sub>s</sub> distribution to homeologs in the related allotetraploid cereal *E. tef* (Figure A3.2) (VanBuren et al., 2020) after conducting mixed modeling on the *K*<sub>s</sub> distributions of both species using wgd v2 (Chen and Zwaenepoel, 2023). All the mixed models, when applied to the *K*<sub>s</sub> of *E. nindensis* V3, had lower peaks (modes) than when applied to *K*<sub>s</sub> of *E. tef* (Figure A3.2). This indicates that *E. nindensis* underwent its whole-genome duplication event much more recently than *E. tef* did, meaning that *E. nindensis* is either an autopolyploid or a recent allopolyploid. In addition, we examined the *k*-mer spectrum of *E. nindensis*, and it follows the pattern that would indicate an autopolyploid (Figure A3.3; see Figure 4A in Becher et al., 2020). This is consistent with our inability to cleanly assemble only two haplotypes for the tetraploid *E.*

*nindensis* genome; autopolyploid genomes are particularly intractable to complete assembly (Sun et al., 2022). Together, these lines of evidence suggest that *E. nindensis* is an autotetraploid.

## DISCUSSION

Availability of quality genome assemblies and annotations vastly expands the research possibilities for plant species, especially wild species, which are under-sequenced compared to crops (Marks et al., 2021). Here, we re-sequenced the genome of *E. nindensis* and greatly improved its contiguity compared to the existing genome version, with only 828 contigs compared to 4,368 in the previous version, and a 20-fold increased contig N50 (Table A3.1) (Pardo et al., 2020). This enabled us to investigate the polyploid origin of *E. nindensis*.

Based on our Ks and *k*-mer evidence, we hypothesize that *E. nindensis* is an autotetraploid. Our GenomeScope *k*-mer spectrum plot (Figure A3.3) shows its highest peak at 1x, indicating many unique *k*-mers, which is characteristic of a high-diversity autotetraploid (Becher et al., 2020). Additionally, given that the peak in the Ks histogram for *E. nindensis* is at a much lower value than that of *E. tef* (Figure A3.2), it is likely that the polyploidy event for *E. nindensis* occurred more recently than that of *E. tef*.

Although the new version of the *E. nindensis* genome presented here is improved, it is still not chromosome scale. In the past, chromosome-scale genomes have been particularly hard to achieve for autopolyploids (Sun et al., 2022), but the autotetraploid rhubarb (Zhang et al., 2024) and autooctoploid sugarcane (Zhang et al., 2018) genomes have both been assembled to chromosome scale, using long-read PacBio HiFi and bacterial artificial chromosome sequencing, respectively. We also used PacBio HiFi to sequence the *E. nindensis* genome, but for improved quality, we should in future scaffold the genome using high-quality HiC chromatin capture data, which was not collected in this study. This could help bring the *E. nindensis* genome closer to chromosome scale.

## REFERENCES

- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al** (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25–29
- Becher H, Brown MR, Powell G, Metherell C, Riddiford NJ, Twyford AD** (2020) Maintenance of Species Differences in Closely Related Tetraploid Parasitic Euphrasia (Orobanchaceae) on an Isolated Island. *Plant Commun* **1**: 100105
- Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J, et al** (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol* **30**: 555–561
- Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, et al** (2004) Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies. *Plant Physiol* **135**: 745–755
- Bray NL, Pimentel H, Melsted P, Pachter L** (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527
- Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al** (2014) MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiol* **164**: 513–524
- Chen H, Zwaenepoel A** (2023) Inference of Ancient Polyploidy from Genomic Data. *In* Y Van de Peer, ed, *Polyploidy Methods Protoc*. Springer US, New York, NY, pp 3–18
- Chen S, Zhou Y, Chen Y, Gu J** (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD** (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J* **89**: 789–804
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H** (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175
- Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H** (2022) Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol* **40**: 1332–1335
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al** (2021) Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008



- De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C** (2018) NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**: 2666–2669
- Eddy SR** (2011) Accelerated Profile HMM Searches. *PLOS Comput Biol* **7**: e1002195
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al** (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al** (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL** (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915
- Korf I** (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59
- Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M** (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**: 278
- Li H** (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100
- Li W, O’Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, et al** (2021) RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Res* **49**: D1020–D1028
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM** (2021) BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**: 4647–4654
- Marçais G, Kingsford C** (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770
- Marks RA, Hotaling S, Frandsen PB, VanBuren R** (2021) Representation and participation across 20 years of plant genome sequencing. *Nat Plants* **7**: 1571–1578
- McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, et al** (2018) The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* **93**: 338–354
- Michael TP, VanBuren R** (2020) Building near-complete plant genomes. *Curr Opin Plant Biol*

- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al** (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**: D412–D419
- Oliver MJ, Tuba Z, Mishler BD** (2000) The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol* **151**: 85–100
- Ou S, Chen J, Jiang N** (2018) Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res* **46**: e126
- Ou S, Jiang N** (2018) LTR\_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant Physiol* **176**: 1410–1422
- Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al** (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**: 275
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, et al** (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* **35**: D883–D887
- Pardo J, Wai CM, Chay H, Madden CF, Hilhorst HWM, Farrant JM, VanBuren R** (2020) Intertwined signatures of desiccation and drought tolerance in grasses. *Proc Natl Acad Sci U S A* **117**: 10079–10088
- Porembski S, Barthlott W** (2000) Granitic and gneissic outcrops (inselbergs) as centers of diversity for desiccation-tolerant vascular plants. *Plant Ecol* **151**: 19–28
- Ranallo-Benavidez TR, Jaron KS, Schatz MC** (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**: 1432
- Roodt R, Spies JJ** (2003) Chromosome studies in the grass subfamily Chloridoideae. II. An analysis of polyploidy. *TAXON* **52**: 736–746
- Stanke M, Diekhans M, Baertsch R, Haussler D** (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644
- Sun Y, Shang L, Zhu Q-H, Fan L, Guo L** (2022) Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* **27**: 391–401
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH** (2008) Synteny and Collinearity in Plant Genomes. *Science* **320**: 486–488
- The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL,**

- Mungall CJ, Basu S, Chisholm RL, Dodson RJ, et al** (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**: D325–D334
- VanBuren R, Wai CM, Keilwagen J, Pardo J** (2018) A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* **2**: e00096
- VanBuren R, Wai CM, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, et al** (2020) Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat Commun* **11**: 1–11
- Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, et al** (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC** (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**: 2202–2204
- Zhang H, He Q, Xing L, Wang R, Wang Y, Liu Y, Zhou Q, Li X, Jia Z, Liu Z, et al** (2024) The haplotype-resolved genome assembly of autotetraploid rhubarb *Rheum officinale* provides insights into its genome evolution and massive accumulation of anthraquinones. *Plant Commun* **5**: 100677
- Zhang H-B, Zhao X, Ding X, Paterson AH, Wing RA** (1995) Preparation of megabase-size DNA from plant nuclei. *Plant J* **7**: 175–184
- Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, et al** (2018) Allele-defined genome of the autopolyloid sugarcane *Saccharum spontaneum* L. *Nat Genet* **50**: 1565–1573

APPENDIX

Table A3.1: Assembly and annotation summary statistics for *Eragrostis nindensis* V3 genome compared to V2.1.

Statistic	V2.1	V3
# of contigs	4,368	828
Contig N50	0.52 Mb	10.43 Mb
Total assembly size	986 Mb	897 Mb
# of gene models (without transposons)	107,683	78,612
BUSCO (complete)	92.10%	92.80%
Mapping rate of single RNA-seq sample (Salmon)	73.60%	76.68%
Whole genome raw LAI	27.42	29.27
Whole genome LAI	21.79	23.64

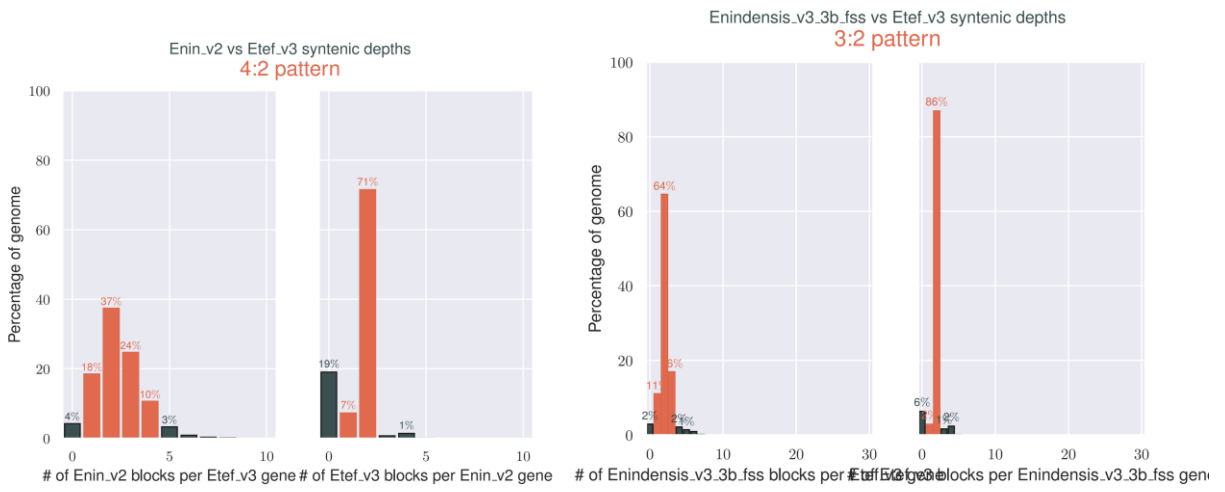


Figure A3.1: Syntenic depth plots for *E. nindensis* version 2 genome (left) and the new version 3 genome (right), each compared to the *E. tef* version 3 genome (VanBuren et al., 2020).

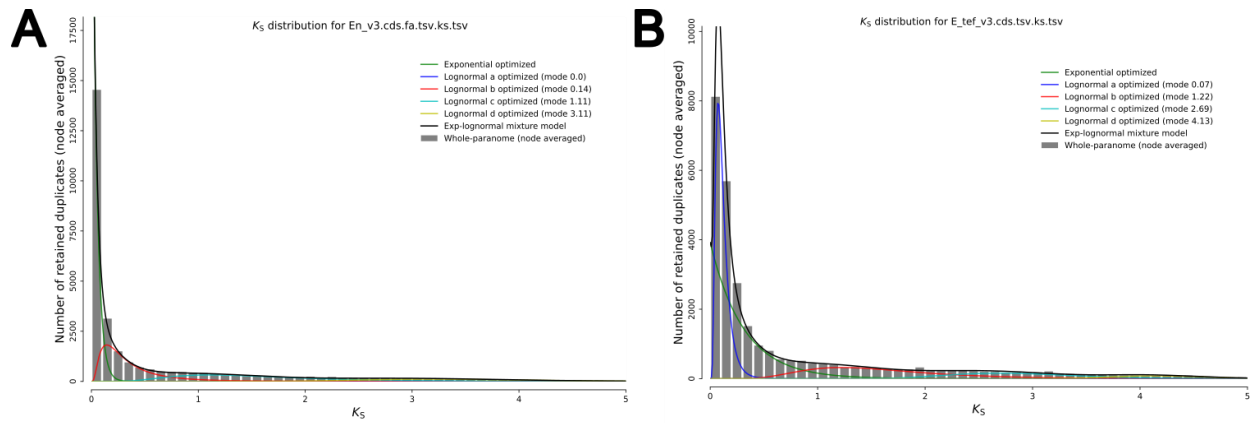


Figure A3.2: Histograms of  $K_s$  (synonymous substitution rate) for the self-comparisons of *E. nindensis* V3 (A) and *E. tef* V3 (B), with fitted mixed models. Note that all models' peaks are left-shifted in *E. nindensis* compared to *E. tef*, i.e. *E. nindensis* has lower  $K_s$  between homeologs than *E. tef* does.

# GenomeScope Profile

len:304,819,709bp uniq:68.4%  
aaaa:95% aaab:3.24% aabb:1.62% aabc:0.001% abcd:0.149%  
kcov:41 err:0.214% dup:1.05 k:21 p:4

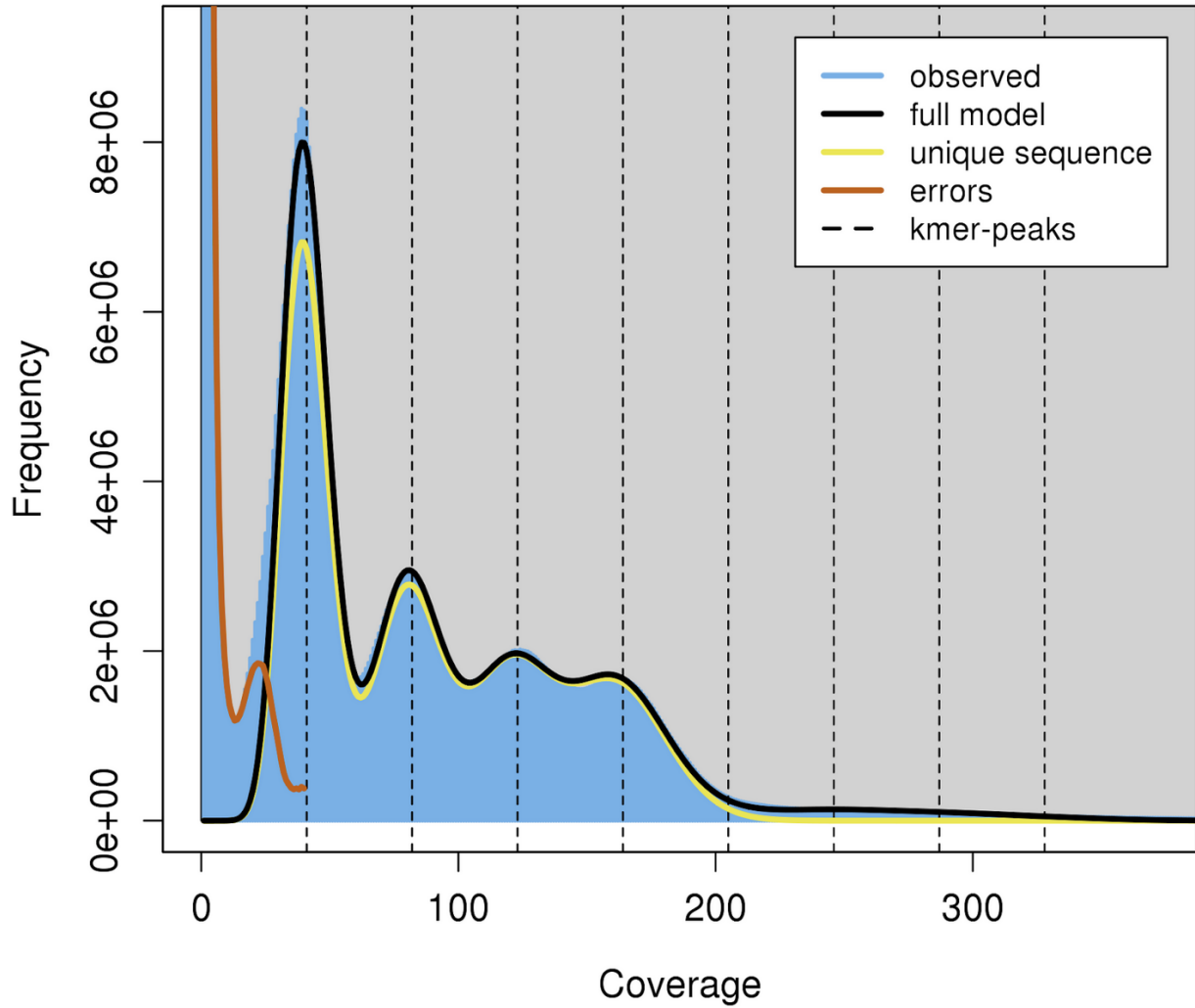


Figure A3.3: GenomeScope *k*-mer spectrum of *E. nindensis* V3. Based on Figure 4A in (Becher et al., 2020), this profile most closely matches that of an autotetraploid.

## **Chapter 4: Genomic signatures of desiccation tolerance in the resilient grass subfamily Chloridoideae**

### **ABSTRACT**

Grasses constitute one of the most ecologically and agriculturally important plant families worldwide. Within the grasses, subfamily Chloridoideae is highly resilient to abiotic stresses, and in particular contains most of the desiccation tolerant grass species. Desiccation tolerance is the ability of vegetative tissues to survive extreme dehydration; its convergent evolution in angiosperms has been linked to both gene family expansion (i.e. early light-induced proteins) and gene expression regulatory changes (i.e. rewiring of seed desiccation tolerance pathways). In this study, we searched for common genomic elements distinguishing desiccation tolerant from sensitive chloridoids, including expanded gene families as well as regulatory motifs in the promoters of differentially expressed genes under desiccation. We found that gene families beyond the previously identified early light-induced proteins were expanded in desiccation tolerant chloridoids, including MYB transcription factors, thaumatin family proteins, and expansin precursors among others. In addition, we found six conserved motifs in differentially expressed genes' promoters which were not found in sensitive species; these motifs are bound by transcription factors of the TIFY, BBR/BPC, and TPC families, which are variously related to development and abiotic stress response. These results indicate that both gene family expansion and gene expression regulation rewiring have contributed to the evolution of desiccation tolerance in the Chloridoideae. In particular, stress-responsive transcription factor families may play a key role.

### **INTRODUCTION**

The grass family, Poaceae, is one of the most ecologically and agronomically important plant families in the world. It contains over 11,700 species in 12 subfamilies, nine of which are contained in two major clades, the BOP (Bambusoideae, Oryzoideae, and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, and Danthonioideae) clades (Huang et al., 2022; Soreng et al., 2022). Chloridoideae is the fourth largest subfamily of the Poaceae, with 1,603 species in 121 genera (Soreng et al., 2022). The three major tribes in the Chloridoideae, namely Cynodonteae, Zoysieae, and Eragrostideae, contain only C4-photosynthetic species (Huang et al., 2022), and Chloridoideae contains many



drought-, salt-, and desiccation-tolerant (DT) grass species (Pardo and VanBuren, 2021), making it a particularly stress-resilient grass subfamily.

Desiccation tolerance is the ability to survive extreme drying. While common in seeds and in the vegetative tissues of non-seed plants, such as bryophytes and ferns, DT of vegetative tissues is uncommon in angiosperms, being present in only 10 angiosperm families (Marks et al., 2021a). Within the Poaceae, nearly all DT species are chloridooids (species in Chloridoideae) (Marks et al., 2021a), and four of these, *Oropetium thomaeum*, *Eragrostis nindensis*, *Sporobolus stapfianus*, and *Tripogon loliiformis*, are among the fifteen most studied DT plant species between 2000 and 2020 (Tebele et al., 2021). Other previously studied DT chloridooids include *Microchloa caffra*, *O. capense*, and *T. minimus*, which are all Cynodonteae species (Marks et al., 2024). There are some differences in the DT physiology and genomics of these species, even though they are in the same subfamily; for example, *E. nindensis* is poikilochlorophyllous (degrades chlorophyll and thylakoids under desiccation) (Pardo et al., 2020), while *O. thomaeum*, *O. capense*, *M. caffra*, and *T. minimus* are homoiochlorophyllous (retains chlorophyll and thylakoids during desiccation) (VanBuren et al., 2017; Marks et al., 2024) and *S. stapfianus* displays characteristics of both strategies (Vecchia et al., 1998). Additionally, while ELIPs are expanded in all DT species with sequenced genomes (VanBuren et al., 2019), they are expanded on different chromosomes in *M. caffra* compared to the *Oropetium* species and *T. minimus*, corresponding to their placement in different subtribes of the Cynodonteae (Marks et al., 2024). In this study, we searched across these differing species to find core genomic elements of DT in the Chloridoideae.

Although vegetative DT was essential for plant colonization of land, it was lost by the angiosperms and re-evolved only in select lineages, likely by rewiring of seed DT pathways (Oliver et al., 2000; Marks et al., 2021a). However, it is likely that the massive expansion of early light-induced proteins (ELIPs) in DT species' genomes (VanBuren et al., 2019) played a role in the evolution of DT as well. In this study, we investigated whether there were other expanded gene families specifically in DT compared to desiccation sensitive (DS) chloridooid grass species. We also investigated changes in gene regulation between DT and DS Chloridoideae species to better understand the evolution of DT in the Chloridoideae.

## METHODS

### *Data gathering and orthogroup construction*

Protein fasta files from the annotation of 67 grass, 16 other monocot, and 4 eudicot species were acquired from public databases, except genomes annotated by the VanBuren lab (Table B4.1). All grass species with easily available proteome fasta files were used. Other monocot species were selected as single representatives of the other 16 sequenced monocot families (outside of Poaceae); the representative species for each family was selected to maximize genome quality and accessibility. Four common model eudicots were included in the analysis as outgroups: *Arabidopsis thaliana*, *Solanum lycopersicum* (tomato), *Medicago truncatula*, and *Vitis vinifera* (grape). Groups of orthologous genes, or orthogroups, were identified for these 87 species using OrthoFinder version 2.5.5 (Emms and Kelly, 2019).

### *Identification of expanded orthogroups*

To identify expanded orthogroups in DT compared to DS chloridoid species, first, orthogroups conserved across the 15 chloridoid species included in the analysis (Figure A4.1) were identified. For each of these orthogroups, the proportion of genes in that orthogroup compared to genes in all conserved orthogroups was calculated in each species. For each orthogroup, the mean proportions for the DS and DT species groups were then calculated, and the difference between DT and DS means was found. A null distribution was generated for all orthogroups by doing 100 permutations of differences of means of random groups of proportions for each orthogroup; the 100 differences of means for each orthogroup were then combined into the grand null distribution. Actual differences of means were compared to the grand null using a 1-sided hypothesis test utilizing the survival function (1-cumulative distribution function). Orthogroups were considered to be significantly expanded in DT compared to DS chloridoids if the p value from the hypothesis test was less than 0.01.

The initial set of 75 expanded orthogroups was subsetted to those meeting one or more of the following conditions: 1) previously found to be expanded in all DT genomes by (VanBuren et al., 2019), i.e. ELIPs, 2) having orthologs or syntenic orthologs (see below for description of synteny analysis) differentially expressed across all 6 DT species, 3) reduced likelihood of being a phylogenetic artifact, by ranking of calculated proportions in all 15 chloridoid species analyzed. Condition 3 was necessary due to the bias toward tribe Cynodonteae, and specifically subtribe Tripogoninae, in the 6 DT species (Figure A4.1). This subset of expanded orthogroups

was characterized by finding the gene name(s) of the *Arabidopsis thaliana* or *Oryza sativa* ortholog(s) present in each orthogroup.

#### *Expression of expanded orthogroups*

Transcripts per million (TPM) expression across control, desiccation/drought, and rehydration time points from pre-existing studies (VanBuren et al., 2017; Pardo et al., 2020; Chávez Montes et al., 2022; Marks et al., 2024) for the 6 DT chloridoid species and two DS species, *Eragrostis tef* and *Sporobolus pyramidalis*, were log transformed and used to evaluate the expression of genes in expanded orthogroups. For each expanded orthogroup, violin plots of log<sub>2</sub> TPM for each gene in the orthogroup, in each time point were created for each species to facilitate cross-species expression comparisons. We also used pre-existing differential expression data to find the percentage of each orthogroup up- and downregulated in each species.

One-to-one syntenic orthologs with *O. thomaeum* as anchor were found using MCScan Python ([https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))) for the other 5 DT chloridoid species and two DS species, *Eragrostis tef* and *Sporobolus pyramidalis*. We then identified sets of syntenic orthologs (syntelogs) conserved across all 8 species. We further identified the conserved syntelogs present in each expanded orthogroup and plotted heatmaps and violin plots of the syntelogs' log<sub>2</sub> TPM for each orthogroup.

#### *Motif analysis*

ATAC-seq reads were predicted from the genomes of each of the 6 DT species and 2 DS species using Predmoter (Kindel et al., 2023) and peaks were called using MACS2 (Zhang et al., 2008). Evaluation of predicted peak locations relative to the transcription start site (TSS) was conducted using deeptools v3.5.1 (Ramírez et al., 2014) plotProfile function. For quality control, predicted data for *O. thomaeum* was compared to pre-existing, real ATAC-seq data for *O. thomaeum* from (St. Aubin et al., 2022); specifically, BEDTools v2.30.0 (Quinlan and Hall, 2010) was used to find overlaps between real and predicted peaks, and plotProfile was also run for the real peaks and compared to the profile for predicted peaks. We found the real and predicted data to be comparable and continued with further analyses using predicted data only.

For motif analysis, our target gene sets for each species consisted of all genes upregulated or downregulated under desiccation/drought compared to control conditions. BEDTools was used to find predicted ATAC-seq peaks falling fully within the promoter regions for these genes, where promoter regions were defined as regions from 0-3 kb upstream of the TSS. HOMER

v4.10 (Heinz et al., 2010) findMotifsGenome.pl was used to find the known and *de novo* motifs enriched in these sets of peaks. We then identified known motifs that were found to be enriched in all six DT species, both those that were also enriched in the two DS species and those that were unique to the DT species.

We used the FIMO functionality of MEME (Bailey et al., 2015) to find which DEGs had each interesting motif in their promoters; upregulated and downregulated genes were considered separately. GO enrichment was run using the weight01 algorithm in topGO (Alexa and Rahnenfuhrer, 2022) for the following sets of genes for each motif: upregulated in DS species, upregulated in DT species, downregulated in DS species, and downregulated in DT species. The resulting sets of enriched GO terms were then compared to find terms unique to DT plants for each motif and set of co-regulated genes.

## RESULTS

### *Orthogroup identification*

From our OrthoFinder run on 84 species (64 grasses, 16 other monocots, and 4 eudicot models), we found 158,026 total orthogroups, of which 8,880 were conserved across the 15 chloridoid grasses included in the analysis, and 7,880 conserved across the chloridoid species and *Arabidopsis thaliana*. Of these, 1,708 were unique to the six DT chloridoid species, and 905 were unique to the nine DS chloridoids.

### *Expanded orthogroups in desiccation-tolerant species*

To find gene families of importance for desiccation tolerance in the Chloridoideae, we ran a statistical test (see Methods for details) to find orthogroups that were expanded in the six DT compared to nine DS species (Figure A4.1). We found 75 expanded orthogroups in this comparison. However, as shown in Figure A4.1, two-thirds of the DT species in this analysis were from the tribe Cynodonteae, meaning that many of these expanded orthogroups were likely phylogenetic artifacts, i.e. they may have been expanded in Cynodonteae only, not truly in all DT species. To remedy this, we examined the rankings of proportions of each orthogroup in all 15 chloridoid species in this analysis. Expanded orthogroups were selected for further analysis if their proportions were higher in 4-6 DT species than most of the 9 DS species. Additionally, if at least one gene in a given orthogroup was differentially expressed in the same direction across all six DT species, that orthogroup was also selected for further analysis. In this way, 17 expanded orthogroups were selected.

For the 17 expanded orthogroups, we identified the *Arabidopsis* orthologs where present or else rice orthologs, and looked up their names in the relevant genome database to gain information about the putative functions of the expanded orthogroups (see Table A4.1). Notable gene families present included the early light-induced proteins (ELIPs), which were included due to their previous identification as important for desiccation tolerance (VanBuren et al., 2019), as well as families newly identified as expanded in DT chloridoids, such as MYB/SANT-like DNA-binding domain proteins in OG0000195 (Table A4.1). The MYB TF family has been previously identified as responsive to various abiotic stresses (Yoon et al., 2020), including desiccation in *Myrothamnus flabellifolia*, a eudicot resurrection plant (Ma et al., 2015), so it makes sense to find this family expanded in DT chloridoids.

Other highly expanded gene families included subtilisin proteases (OG0000047), pathogenesis-related thaumatin superfamily proteins (OG0000067), and putative, expressed expansin precursors (OG0000164; Table A4.1). Subtilisin proteases, or subtilases, have been found to be involved in both biotic and abiotic stress in the past (Figueiredo et al., 2018). Overexpression of thaumatin-like proteins has been found to improve drought and salt tolerance in broccoli (He et al., 2021) and drought tolerance in *Arabidopsis* (Muoki et al., 2021). Furthermore, thaumatins were among the proteins whose abundance increased under desiccation in the DT eudicot species *Haberlea rhodopensis* (Mladenov et al., 2022). Our finding that thaumatin superfamily proteins are expanded in DT chloridoids thus adds to pre-existing evidence of their importance for desiccation response.

Desiccation includes the physical loss of water from cells, which can cause great stress for cell walls (Moore et al., 2008). Expansins, which act to increase the flexibility of the cell wall (Marowa et al., 2016), are more active under desiccation and rehydration in the eudicot DT plant *Craterostigma plantagineum* (Jones and McQueen-Mason, 2004). In this study, we found that one of the expanded orthogroups in DT chloridoids contains putative expansin precursors, indicating the key role of expansins in DT in the Chloridoideae.

#### *Expression of expanded orthogroups*

We evaluated the expression of the expanded orthogroups using heatmaps and violin plots of the expression (log<sub>2</sub> TPM) of syntelogs within each expanded orthogroup, if any, that were conserved across all eight species for which expression data was present. There were no conserved syntelogs present in OG0000195, OG0000050, OG0002131, or OG0000418, so they

were excluded from this analysis. All other expanded orthogroups had between 2 and 11 conserved syntelogs. As expected, the syntelogs in each orthogroup did not share the same expression profiles across species and conditions. OG0000033, which contains protein kinases, is a good example of this. In OG0000033, most of the 7 conserved syntelogs are similarly expressed between DT and DS species under control and dehydration conditions, but one, Syn6286, is more highly expressed in DT than DS species under both conditions (Figure A4.2). In contrast, of the three conserved syntelogs in OG0000047 (subtilisin proteases), one of them has higher mean expression in control and dehydration conditions for DS species. Other orthogroups, such as OG0000055 (protein kinases), OG0000164 (expansin precursors), OG0000240 (CCR-associated factor deadenylases), and OG0000355 (cyclin T1) had some syntelogs with higher expression in DT species and others with higher expression in DS species.

Notably, in OG0000413 (ELIPs), all 8 conserved syntelogs had higher expression in DS species for at least one condition (Figure A4.3). ELIPs have been previously reported as highly expanded in DT plant species (VanBuren et al., 2019; Marks et al., 2024), as confirmed by this study. Overall, we found that ELIP expression, both for conserved and non-conserved copies, increases under dehydration, in both DT and DS species (Figure B4.1). Thus, it is possible that the upregulated ELIP copies under desiccation in DT chloridoids are copies that are unique to DT species.

#### *Cis element identification*

Desiccation tolerance, although key for plants' movement onto land, was subsequently lost in the angiosperms and convergently re-evolved in vegetative tissues of certain lineages (Oliver et al., 2000; VanBuren et al., 2019). The predominant hypothesis is that vegetative DT evolved by rewiring, i.e. changing gene expression regulation, of the seed DT pathways present in most angiosperms (Oliver et al., 2000). However, gene duplication can also contribute to evolution of new traits, for example ELIPs (VanBuren et al., 2019) and other expanded gene families (Table A4.1) in DT genomes. To address the possibility of gene regulatory changes contributing to evolution of DT in Chloridoideae, here we searched for common cis regulatory elements across species in genes differentially expressed under desiccation to see if the same transcription factor networks have been rewired to function in DT across DT chloridoids.

To identify cis regulatory elements, we first predicted open chromatin regions in the six DT and two DS chloridoids for which we had expression data (Table B4.2), using Predmoter

(Kindel et al., 2023), followed by motif analysis in the predicted ATAC-seq peak regions found in DEGs' promoters using HOMER (Heinz et al., 2010). This enabled us to reduce false positive motifs, i.e. enriched motifs present in closed chromatin regions. For quality control, we compared the peaks predicted by Predmoter for *O. thomaeum* with real *O. thomaeum* ATAC-seq peaks from (St. Aubin et al., 2022). We found that, although there were drastically fewer real peaks (26,321) than predicted peaks (143,049), 98% of real peaks overlapped with predicted peaks; thus, we proceeded to use the predicted data for downstream analysis. The species with the most predicted peaks was *S. stapfianus* with 420,639 peaks; the species with the fewest was *T. minimus* with 115,084 (Table B4.3). In general, polyploid species had more predicted peaks than diploids (Table B4.3). We further did quality control of these peaks with deepTools plotProfile (Ramírez et al., 2014) and found that most predicted peaks were found just before the transcription start site (TSS) of genes (Figure B4.2), except in the two *Eragrostis* species, for which there were more peaks after than before the TSS. However, even for the *Eragrostis* species, there is still a substantial dip in peak frequency at the TSS, as expected.

For all species except *E. nindensis*, we used previously calculated differentially expressed genes (DEGs; see Table B4.2 for citations) from previous studies for downstream cis element analysis. For *E. nindensis*, we re-calculated DEGs from data re-analyzed using the improved version of the genome, V3 (see Chapter 3 of this dissertation). Table B4.4 shows the numbers of overall DEGs and upregulated and downregulated genes found for desiccation vs. control contrasts in each species. Numbers of DEGs ranged from 9,520 overall in *E. tef* to 43,794 overall in *S. stapfianus* (Table B4.4). These numbers included only desiccation time points.

We used HOMER to find enriched motifs in the promoters of DEGs for the 6 DT Chloridoideae species (*O. thomaeum*, *O. capense*, *E. nindensis*, *M. caffra*, *T. minimus*, *S. stapfianus*) and 2 DS sister species (*E. tef* and *S. pyramidalis*). Motifs were found separately for up- and downregulated genes in each species. Table 2 shows the motifs and associated transcription factors that were enriched across DEGs' promoters for all 8 species, as well as those that were enriched in DT species but not in DS species (Figure A4.4), which are of particular interest. The motif bound by the ZIM transcription factor, which is part of the TIFY family (Singh and Mukhopadhyay, 2021), was enriched for both up- and downregulated gene sets in DT species only. This TF family has been found to be involved in regulation of responses to various abiotic stressors in wheat (Singh and Mukhopadhyay, 2021). The TIFY family has

also previously been found to be upregulated during desiccation in the DT bryophyte *Bryum argenteum* (Gao et al., 2017), but it has not previously been linked to regulation of other desiccation-responsive genes to our knowledge.

Following the motif enrichment, we used MEME FIMO to find the genes associated with each motif in the genomes of each of the eight species and ran GO term enrichment on these genes using topGO (Alexa and Rahnenfuhrer, 2022). The only GO terms enriched in DEG sets associated with the ZIM motif were “translation” and “DNA-templated transcription termination”, indicating possible regulatory functions for these genes. For the ZML1 motif, no enriched GO terms were found for upregulated genes from any species set or for downregulated genes from DS species; the only GO term enriched in downregulated genes from DT species was “chloroplast rRNA processing.”

By far the motif with the largest number of enriched GO terms in its associated downregulated genes was BPC6, a member of the BARLEY B-RECOMBINANT/BASIC PENTACYSTEINE (BBR/BPC) TF family. This family acts to repress target genes via recruitment of the polycomb repressive complex (Sahu et al., 2023), which is likely why we found the BPC6 and BPC1 motifs to be enriched only in downregulated genes’ promoters (Table A4.2). Various abiotic stress response-related GO terms were enriched in downregulated genes associated with BPC6, in DT species only; this was also the case for various regulation-related GO terms. It has previously been found that when the BPC1 and BPC2 genes are knocked out in *Arabidopsis thaliana*, salt tolerance is reduced (Sahu et al., 2023). Our results suggest that this TF family is also important for downregulation of genes in DT species.

The At2g45680(TCP) motif had no enriched GO terms for its associated genes in DS species. In DT species, the only enriched GO term for its associated downregulated genes was “uroporphyrinogen III biosynthetic process”, which is likely related to chlorophyll biosynthesis since uroporphyrinogen III is a precursor of chlorophyll. Whether they are homoiochlorophyllous (retain chlorophyll and thylakoids under desiccation) or poikilochlorophyllous (degrade chlorophyll and thylakoids), it makes sense that DT species would inhibit further chlorophyll biosynthesis. This result indicates that the TCP TF family, specifically DT plants’ orthologs of At2g45680 (TCP9), may be involved in regulation of this key desiccation response in the DT species in the Chloridoideae.



## DISCUSSION

With an increase in drought events expected globally due to climate change, it is essential to understand plants' natural adaptations to extreme dryness, including vegetative desiccation tolerance, in order to gain insights that will further crop improvement. Within the economically important grass family, Poaceae, many of the DT species are found in the highly resilient subfamily Chloridoideae (Marks et al., 2021a). In this study, we used comparative genomics of 6 DT (*Eragrostis nindensis*, *Oropetium thomaeum*, *O. capense*, *Microchloa caffra*, *Tripogon minimus*, *Sporobolus stapfianus*) and 9 DS chloridoid grass species to find expanded gene families in DT chloridoids. Further, we used expression data from the same 6 DT and 2 DS (*E. tef*, *S. pyramidalis*) chloridoids to study changes in gene regulation between DT and DS chloridoids using motif analysis. We found 17 expanded orthogroups, including ELIPs which have previously been identified as expanded in all DT genomes (VanBuren et al., 2019) as well as other, newly discovered expanded gene families (Table A4.1). We also found six motifs (Figure A4.4) enriched in the promoters of dehydration-responsive genes in DT species but not in DS species.

The early light-induced proteins (ELIPs) have previously been found to be expanded in all DT plants' genomes; given the high light and oxidative stresses associated with desiccation, it is likely that these proteins function in cellular photoprotection, essential for DT (VanBuren et al., 2019). We found that overall, the expression of ELIPs increases during dehydration stress regardless of the tolerance or sensitivity of the species (Figure B4.1). However, ELIP syntenic orthologs conserved across the 8 species were generally expressed more highly in DS than DT species, both during control and dehydration (Figure A4.3B).

We also found other gene families beyond ELIPs that were expanded in DT chloridoid species. Some of the most significantly expanded included MYB/SANT-like DNA-binding domain proteins (i.e. MYB family transcription factors), subtilisin proteases, expansin precursors, and thaumatin superfamily proteins (Table A4.1). All of these gene families have previously been related to abiotic stress in general, and some to desiccation. For example, the MYB transcription factor family has been found to be involved in response to multiple abiotic stresses, including drought (Yoon et al., 2020; Wang et al., 2021b). MYBs are known to regulate stomatal closure and movement through ABA-mediated pathways during drought stress (Wang et al., 2021b). Furthermore, in DT eudicot *Myrothamnus flabellifolia*, the MYB family was

identified as desiccation-responsive (Ma et al., 2015). However, this is the first evidence that the MYB family is expanded in any set of DT species. It could be suggested that MYB transcription factors are important regulators of desiccation-responsive genes in DT chloridoids; however, we did not find any MYB motifs to be enriched in DEGs across all six DT species in this analysis (Figure A4.4). Thus, different members of the MYB family may be active in different DT species, making the family as a whole important for DT.

Under drought, the subtilisin proteases, also known as subtilases, cleave other proteins to produce small peptides which are involved in signaling pathways such as the mitogen-activated protein kinase (MAPK) and calcium pathways (Datta et al., 2024). In addition, proteases of this class were active during desiccation in the DT eudicot *Ramonda serbica* (Kidrič et al., 2014), indicating that they may be connected to DT in this species. To our knowledge, our finding that subtilase genes are expanded in DT chloridoids is the first indication of these proteins' importance in DT monocots, and this gene family's role in desiccation should be studied further.

We also found that expansin precursor genes were expanded in DT chloridoids compared to DS chloridoids. Desiccation causes many physical stresses for cells, including cell wall deformation (Moore et al., 2008). Expansins help increase cell wall elasticity, which is essential for desiccation tolerance (Jones and McQueen-Mason, 2004; Moore et al., 2008). That being said, this has mainly been studied in DT eudicots such as *Craterostigma plantagineum* (Jones and McQueen-Mason, 2004), whose cell wall composition differs from that of the grasses (Neeragunda Shivaraj et al., 2018); therefore, we recommend further study of the role of expansins in DT of monocots, particularly grasses.

Thaumatococcus superfamily proteins are known to be responsive to both pathogen-induced and abiotic stresses, including stressors with osmotic components such as drought and salt (Liu et al., 2010), but to our knowledge, they have not been studied in DT plants specifically. Here, we found that thaumatococcus superfamily proteins constituted one of the most significantly expanded gene families in DT chloridoid grasses (Table A4.1). We suggest, based on their roles in signaling during pathogen infection by binding to glycoproteins (Liu et al., 2010), that thaumatococcus superfamily proteins may have a similar role in desiccation response in DT plants. This should be studied further.

In this study, we also found a set of TF binding motifs enriched in the promoters of desiccation-responsive genes across all six DT chloridoids, which were also not found in DS

chloridoids' dehydration-responsive gene promoters (Figure A4.4). TF families represented here included TIFY (ZIM and ZML1), BBR/BPC (BPC1 and BPC6), and TCP (TCP16 and At2g45680/TCP9). Most of these were associated with promoters of genes downregulated under desiccation, although the ZIM motif was also associated with upregulated genes (Table A4.2). TIFY genes responsive to various hormone treatments as well as abiotic and biotic stresses have been identified in various species, including wheat as well as other crops (Chini et al., 2017; He et al., 2020; Wang et al., 2020a; Zhang et al., 2020; Singh and Mukhopadhyay, 2021; Liu et al., 2022a; Liu et al., 2022b; Zhao et al., 2023), but to our knowledge have been less studied in wild plants and not at all in DT species. Our findings suggest that TIFY TFs, in particular the ZML subfamily, are important regulators of desiccation-responsive genes in DT grasses and should be investigated more in future.

The BARLEY B-RECOMBINANT/BASIC PENTACYSTEINE (BBR/BPC) TF family is conserved across plants and its members generally function as transcriptional repressors, via recruitment of the polycomb repressive complex (Sahu et al., 2023). Consistent with this, BPC1, which is one of the BBR/BPC motifs we identified as enriched in downregulated genes across DT chloridoids, has been found to improve salt tolerance by repressing *GALACTAN SYNTHASE 1*, a biosynthetic gene for a cell wall galactan that leads to salt hypersensitivity response (Yan et al., 2021a). In addition, BPC6, the TF binding to the other important motif we identified, was found to be a key regulator of cuticular wax biosynthesis (Sahu et al., 2023). To our knowledge, we are the first to identify these TFs as key desiccation regulators in DT grasses.

The final family we identified as key DT regulators was the TEOSINTE BRANCHED/CYCLOIDEA/PROLIFERATING CELL FACTOR (TCP) family, a subset of the basic helix-loop-helix (bHLH) family. These proteins are most known for developmental regulation (Li, 2015). Specifically, TCP16 has been linked to early pollen development (Takeda et al., 2006) and diel regulation of copper transport (Andrés-Colás et al., 2018). TCP9 has been found to regulate root system developmental plasticity during pathogen infection, including oxidative stress response genes (Willig et al., 2022). It is important to note that here, we have identified key motifs, including these TCP motifs, based on expression from leaf tissues only; given that TCP9 has been linked to root development in the past, it may be worth investigating the role of this TF in roots of DT plants. However, it is also clear that the TCP TFs are important

regulators of desiccation-downregulated genes in leaves, and may be linked to developmental changes under desiccation. Further study is required to confirm this.

In sum, we have identified a number of important gene families and regulators for desiccation tolerance in the highly resilient grass subfamily Chloridoideae. We found several novel gene families of importance to DT, in addition to confirming the importance of the highly expanded ELIPs. We also identified six motifs, bound by proteins from three transcription factor families, which are enriched in desiccation-responsive (mainly downregulated) genes' promoters in all six DT species studied here, but not in DS species. Therefore, we can conclude that DT has evolved via both changes in gene regulation and expansion of certain gene families in the Chloridoideae. We also recommend further investigation into these gene families and TFs as potential targets for improvement of drought tolerance in crops, as well as for understanding of DT in the economically important grass family.

## REFERENCES

- Akpınar BA, Biyiklioglu S, Alptekin B, Havránková M, Vrána J, Doležel J, Distelfeld A, Hernandez P, Iwgc T, Budak H** (2018) Chromosome-based survey sequencing reveals the genome organization of wild wheat progenitor *Triticum dicoccoides*. *Plant Biotechnol J* **16**: 2077–2087
- Alexa A, Rahnenfuhrer J** (2022) topGO: Enrichment Analysis for Gene Ontology.
- Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, Liu J, Pan L, Zhang T, et al** (2013) Genome sequence of the date palm *Phoenix dactylifera* L. *Nat Commun* **4**: 2274
- Andrés-Colás N, Carrió-Seguí A, Abdel-Ghany SE, Pilon M, Peñarrubia L** (2018) Expression of the Intracellular COPT3-Mediated Cu Transport Is Temporally Regulated by the TCP16 Transcription Factor. *Front Plant Sci*. doi: 10.3389/fpls.2018.00910
- Bailey TL, Johnson J, Grant CE, Noble WS** (2015) The MEME Suite. *Nucleic Acids Res* **43**: W39–W49
- Bayer PE, Fraser MW, Martin BC, Petereit J, Severn-Ellis AA, Sinclair EA, Batley J, Kendrick GA, Edwards D** (2022) Not all pathways are the same – unique adaptations to submerged environments emerge from comparative seagrass genomics. 2022.11.22.517588
- Beier S, Himmelbach A, Colmsee C, Zhang X-Q, Barrero RA, Zhang Q, Li L, Bayer M, Bolser D, Taudien S, et al** (2017) Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. *Sci Data* **4**: 170044
- Bredeson JV, Lyons JB, Oniyinde IO, Okereke NR, Kolade O, Nnabue I, Nwadike CO, Hřibová E, Parker M, Nwogha J, et al** (2022) Chromosome evolution and the genetic basis of agronomically important traits in greater yam. *Nat Commun* **13**: 2001
- Bruccoleri RE, Oakeley EJ, Faust AME, Altorfer M, Dessus-Babus S, Burckhardt D, Oertli M, Naumann U, Petersen F, Wong J** (2023) Genome assembly of the bearded iris *Iris pallida* Lam. 2023.08.29.555454
- Carballo J, Santos B a. CM, Zappacosta D, Garbus I, Selva JP, Gallo CA, Díaz A, Albertini E, Caccamo M, Echenique V** (2019) A high-quality genome of *Eragrostis curvula* grass provides insights into Poaceae evolution and supports new strategies to enhance forage quality. *Sci Rep* **9**: 10250
- Chávez Montes RA, Haber A, Pardo J, Powell RF, Divisetty UK, Silva AT, Hernández-Hernández T, Silveira V, Tang H, Lyons E, et al** (2022) A comparative genomics examination of desiccation tolerance and sensitivity in two sister grass species. *Proc Natl Acad Sci* **119**: e2118886119

- Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis Goicoechea J, Liang C, et al** (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun* **4**: 1595
- Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD** (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* **89**: 789–804
- Chini A, Ben-Romdhane W, Hassairi A, Aboul-Soud MAM** (2017) Identification of TIFY/JAZ family genes in *Solanum lycopersicum* and their regulation in response to abiotic stresses. *PLOS ONE* **12**: e0177381
- Costa M-CD, Artur MAS, Maia J, Jonkheer E, Derks MFL, Nijveen H, Williams B, Mundree SG, Jiménez-Gómez JM, Hesselink T, et al** (2017) A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nat Plants* **3**: 1–10
- Datta T, Kumar RS, Sinha H, Trivedi PK** (2024) Small but mighty: Peptides regulating abiotic stress responses in plants. *Plant Cell Environ* **47**: 1207–1223
- D’Hont A, Denoeud F, Aury J-M, Baurens F-C, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al** (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**: 213–217
- Emms DM, Kelly S** (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**: 238
- Figueiredo J, Sousa Silva M, Figueiredo A** (2018) Subtilisin-like proteases in plant defence: the past, the present and beyond. *Mol Plant Pathol* **19**: 1017–1028
- Finkers R, van Kaauwen M, Ament K, Burger-Meijer K, Egging R, Huits H, Kodde L, Kroon L, Shigyo M, Sato S, et al** (2021) Insights from the first genome assembly of Onion (*Allium cepa*). *G3 GenesGenomesGenetics* **11**: jkab243
- Gao B, Li X, Zhang D, Liang Y, Yang H, Chen M, Zhang Y, Zhang J, Wood AJ** (2017) Desiccation tolerance in bryophytes: The dehydration and rehydration transcriptomes in the desiccation-tolerant bryophyte *Bryum argenteum*. *Sci Rep* **7**: 7571
- Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session A, Martin J, Cartwright A, Katz A, et al** (2020) Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat Commun* **11**: 3670
- Guo L, Qiu J, Ye C, Jin G, Mao L, Zhang H, Yang X, Peng Q, Wang Y, Jia L, et al** (2017) *Echinochloa crus-galli* genome analysis provides insight into its adaptation and invasiveness as a weed. *Nat Commun* **8**: 1031

- Guo Z-H, Ma P-F, Yang G-Q, Hu J-Y, Liu Y-L, Xia E-H, Zhong M-C, Zhao L, Sun G-L, Xu Y-X, et al** (2019) Genome Sequences Provide Insights into the Reticulate Origin and Unique Traits of Woody Bamboos. *Mol Plant* **12**: 1353–1365
- Haas M, Kono T, Macchietto M, Millas R, McGilp L, Shao M, Duquette J, Qiu Y, Hirsch CN, Kimball J** (2021) Whole-genome assembly and annotation of northern wild rice, *Zizania palustris* L., supports a whole-genome duplication in the *Zizania* genus. *Plant J* **107**: 1802–1818
- Han B, Jing Y, Dai J, Zheng T, Gu F, Zhao Q, Zhu F, Song X, Deng H, Wei P, et al** (2020) A Chromosome-Level Genome Assembly of *Dendrobium huoshanense* Using Long Reads and Hi-C Data. *Genome Biol Evol* **12**: 2486–2490
- Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, et al** (2017) The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun* **8**: 1279
- He L, Li L, Zhu Y, Pan Y, Zhang X, Han X, Li M, Chen C, Li H, Wang C** (2021) *BolTLP1*, a Thaumatin-like Protein Gene, Confers Tolerance to Salt and Drought Stresses in Broccoli (*Brassica oleracea* L. var. *italica*). *Int J Mol Sci* **22**: 11132
- He X, Kang Y, Li W, Liu W, Xie P, Liao L, Huang L, Yao M, Qian L, Liu Z, et al** (2020) Genome-wide identification and functional analysis of the TIFY gene family in the response to multiple stresses in *Brassica napus* L. *BMC Genomics* **21**: 736
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK** (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589
- Hittalmani S, Mahesh HB, Shirke MD, Biradar H, Uday G, Aruna YR, Lohithaswa HC, Mohanrao A** (2017) Genome and Transcriptome sequence of Finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics* **18**: 465
- Hofstatter PG, Thangavel G, Lux T, Neumann P, Vondrak T, Novak P, Zhang M, Costa L, Castellani M, Scott A, et al** (2022) Repeat-based holocentromeres influence genome architecture and karyotype evolution. *Cell* **185**: 3153-3168.e18
- Huang W, Zhang L, Columbus JT, Hu Y, Zhao Y, Tang L, Guo Z, Chen W, McKain M, Bartlett M, et al** (2022) A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C4 photosynthesis. *Mol Plant* **15**: 755–777
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al** (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662

- Jones L, McQueen-Mason S** (2004) A role for expansins in dehydration and rehydration of the resurrection plant *Craterostigma plantagineum*. *FEBS Lett* **559**: 61–65
- Kamal N, Tsardakas Renhuldt N, Bentzer J, Gundlach H, Haberer G, Juhász A, Lux T, Bose U, Tye-Din JA, Lang D, et al** (2022) The mosaic oat genome gives insights into a uniquely healthy cereal crop. *Nature* **606**: 113–119
- Kidrič M, Sabotič J, Stevanović B** (2014) Desiccation tolerance of the resurrection plant *Ramonda serbica* is associated with dehydration-dependent changes in levels of proteolytic activities. *J Plant Physiol* **171**: 998–1002
- Kindel F, Triesch S, Schlüter U, Randarevitch LA, Reichel-Deland V, Weber APM, Denton AK** (2023) Predmoter - Cross-species prediction of plant promoter and enhancer regions. 2023.11.03.565452
- Knorst V, Yates S, Byrne S, Asp T, Widmer F, Studer B, Kölliker R** (2019) First assembly of the gene-space of *Lolium multiflorum* and comparison to other Poaceae genomes. *Grassl Sci* **65**: 125–134
- Kuang L, Shen Q, Chen L, Ye L, Yan T, Chen Z-H, Waugh R, Li Q, Huang L, Cai S, et al** (2022) The genome and gene editing system of sea barleygrass provide a novel platform for cereal domestication and stress tolerance studies. *Plant Commun* **3**: 100333
- Li G, Wang L, Yang J, He H, Jin H, Li X, Ren T, Ren Z, Li F, Han X, et al** (2021a) A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat Genet* **53**: 574–584
- Li H-L, Wu L, Dong Z, Jiang Y, Jiang S, Xing H, Li Q, Liu G, Tian S, Wu Z, et al** (2021b) Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic Res* **8**: 1–13
- Li S** (2015) The *Arabidopsis thaliana* TCP transcription factors: A broadening horizon beyond development. *Plant Signal. Behav.*
- Li W, Shi C, Li K, Zhang Q-J, Tong Y, Zhang Y, Wang J, Clark L, Gao L-Z** (2021c) Draft genome of the herbaceous bamboo *Raddia distichophylla*. *G3 GenesGenomesGenetics* **11**: jkaa049
- Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, et al** (2018) Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **557**: 424–428
- Liu J-J, Sturrock R, Ekramoddoullah AKM** (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Rep* **29**: 419–436



- Liu X, Yu F, Yang G, Liu X, Peng S** (2022a) Identification of TIFY gene family in walnut and analysis of its expression under abiotic stresses. *BMC Genomics* **23**: 190
- Liu Y-L, Zheng L, Jin L-G, Liu Y-X, Kong Y-N, Wang Y-X, Yu T-F, Chen J, Zhou Y-B, Chen M, et al** (2022b) Genome-Wide Analysis of the Soybean TIFY Family and Identification of GmTIFY10e and GmTIFY10g Response to Salt Stress. *Front Plant Sci.* doi: 10.3389/fpls.2022.845314
- Lovell JT, Jenkins J, Lowry DB, Mamidi S, Sreedasyam A, Weng X, Barry K, Bonnette J, Campitelli B, Daum C, et al** (2018) The genomic landscape of molecular responses to natural drought stress in *Panicum hallii*. *Nat Commun* **9**: 5213
- Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al** (2021) Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* **590**: 438–444
- Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, et al** (2017) Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature* **551**: 498–502
- Ma C, Wang H, Macnish AJ, Estrada-Melo AC, Lin J, Chang Y, Reid MS, Jiang C-Z** (2015) Transcriptomic analysis reveals numerous diverse protein kinases and transcription factors involved in desiccation tolerance in the resurrection plant *Myrothamnus flabellifolia*. *Hortic Res.* doi: 10.1038/hortres.2015.34
- Ma P-F, Liu Y-L, Jin G-H, Liu J-X, Wu H, He J, Guo Z-H, Li D-Z** (2021) The *Pharus latifolius* genome bridges the gap of early grass evolution. *Plant Cell* **33**: 846–864
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, et al** (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* **51**: 885–895
- Mamidi S, Healey A, Huang P, Grimwood J, Jenkins J, Barry K, Sreedasyam A, Shu S, Lovell JT, Feldman M, et al** (2020) A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. *Nat Biotechnol* **38**: 1203–1210
- Marks RA, Farrant JM, Nicholas McLetchie D, VanBuren R** (2021) Unexplored dimensions of variability in vegetative desiccation tolerance. *Am J Bot* **108**: 346–358
- Marks RA, Pas LVD, Schuster J, Gilman IS, VanBuren R** (2024) Convergent evolution of desiccation tolerance in grasses. 2023.11.29.569285
- Marowa P, Ding A, Kong Y** (2016) Expansins: roles in plant growth and potential applications in crop improvement. *Plant Cell Rep* **35**: 949–965

- McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B, et al** (2018) The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* **93**: 338–354
- Miao J, Feng Q, Li Y, Zhao Q, Zhou C, Lu H, Fan D, Yan J, Lu Y, Tian Q, et al** (2021) Chromosome-scale assembly and analysis of biomass crop *Miscanthus lutarioriparius* genome. *Nat Commun* **12**: 2458
- Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, et al** (2015) The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* **47**: 1435–1442
- Mitros T, Session AM, James BT, Wu GA, Belaffif MB, Clark LV, Shu S, Dong H, Barling A, Holmes JR, et al** (2020) Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun* **11**: 5442
- Mladenov P, Zasheva D, Planchon S, Leclercq CC, Falconet D, Moyet L, Brugière S, Moyankova D, Tchorbadjieva M, Ferro M, et al** (2022) Proteomics Evidence of a Systemic Response to Desiccation in the Resurrection Plant *Haberlea rhodopensis*. *Int J Mol Sci* **23**: 8520
- Moore JP, Vicré-Gibouin M, Farrant JM, Driouich A** (2008) Adaptations of higher plant cell walls to water loss: drought vs desiccation. *Physiol Plant* **134**: 237–245
- Mueller LA, Lankhorst RK, Tanksley SD, Giovannoni JJ, White R, Vrebalov J, Fei Z, van Eck J, Buels R, Mills AA, et al** (2009) A Snapshot of the Emerging Tomato Genome Sequence. *Plant Genome*. doi: 10.3835/plantgenome2008.08.0005
- Muoki RC, Paul A, Kaachra A, Kumar S** (2021) Membrane localized *thaumatin-like protein* from tea (*CsTLP*) enhanced seed yield and the plant survival under drought stress in *Arabidopsis thaliana*. *Plant Physiol Biochem* **163**: 36–44
- Nagy I, Veeckman E, Liu C, Bel MV, Vandepoele K, Jensen CS, Ruttink T, Asp T** (2022) Chromosome-scale assembly and annotation of the perennial ryegrass genome. *BMC Genomics* **23**: 1–20
- Neeragunda Shivaraj Y, Barbara P, Gugi B, Vicré-Gibouin M, Driouich A, Ramasandra Govind S, Devaraja A, Kambalagere Y** (2018) Perspectives on Structural, Physiological, Cellular, and Molecular Responses to Desiccation in Resurrection Plants. *Scientifica* **2018**: e9464592
- Oliver MJ, Tuba Z, Mishler BD** (2000) The evolution of vegetative desiccation tolerance in land plants. *Plant Ecol* **151**: 85–100

- Olsen JL, Rouzé P, Verhelst B, Lin Y-C, Bayer T, Collen J, Dattolo E, De Paoli E, Dittami S, Maumus F, et al** (2016) The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. *Nature* **530**: 331–335
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, et al** (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res* **35**: D883–D887
- Pardo J, Man Wai C, Chay H, Madden CF, Hilhorst HWM, Farrant JM, VanBuren R** (2020) Intertwined signatures of desiccation and drought tolerance in grasses. *Proc Natl Acad Sci* **117**: 10079–10088
- Pardo J, VanBuren R** (2021) Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *Plant Cell* **33**: 3391–3401
- Paril J, Pandey G, Barnett EM, Rane RV, Court L, Walsh T, Fournier-Level A** (2022) Rounding up the annual ryegrass genome: High-quality reference genome of *Lolium rigidum*. *Front. Genet.* 13:
- Planta J, Liang Y-Y, Xin H, Chansler MT, Prather LA, Jiang N, Jiang J, Childs KL** (2022) Chromosome-scale genome assemblies and annotations for Poales species *Carex cristatella*, *Carex scoparia*, *Juncus effusus*, and *Juncus inflexus*. *G3 GenesGenomesGenetics* **12**: jkac211
- Quinlan AR, Hall IM** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842
- Ramakrishnan M, Yrjälä K, Vinod KK, Sharma A, Cho J, Satheesh V, Zhou M** (2020) Genetics and genomics of moso bamboo (*Phyllostachys edulis*): Current status, future challenges, and biotechnological opportunities toward a sustainable bamboo industry. *Food Energy Secur* **9**: e229
- Ramírez F, DüNDAR F, Diehl S, Grüning BA, Manke T** (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187–W191
- Reuscher S, Furuta T, Bessho-Uehara K, Cosi M, Jena KK, Toyoda A, Fujiyama A, Kurata N, Ashikari M** (2018) Assembling the genome of the African wild rice *Oryza longistaminata* by exploiting synteny in closely related *Oryza* species. *Commun Biol* **1**: 1–10
- Sahu A, Singh R, Verma PK** (2023) Plant BBR/BPC transcription factors: unlocking multilayered regulation in development, stress and immunity. *Planta* **258**: 31
- Seetharam AS, Yu Y, Bélanger S, Clark LG, Meyers BC, Kellogg EA, Hufford MB** (2021) The *Streptochoaeta* Genome and the Evolution of the Grasses. *Front. Plant Sci.* 12:

- Singh P, Mukhopadhyay K** (2021) Comprehensive molecular dissection of TIFY Transcription factors reveal their dynamic responses to biotic and abiotic stress in wheat (*Triticum aestivum* L.). *Sci Rep* **11**: 9739
- Soreng RJ, Peterson PM, Zuloaga FO, Romaschenko K, Clark LG, Teisher JK, Gillespie LJ, Barberá P, Welker CAD, Kellogg EA, et al** (2022) A worldwide phylogenetic classification of the Poaceae (Gramineae) III: An update. *J Syst Evol* **60**: 476–521
- St. Aubin B, Wai CM, Kenchanmane Raju SK, Niederhuth CE, VanBuren R** (2022) Regulatory dynamics distinguishing desiccation tolerance strategies within resurrection grasses. *Plant Direct* **6**: e457
- Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL, et al** (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* **50**: 285–296
- Studer AJ, Schnable JC, Weissmann S, Kolbe AR, McKain MR, Shao Y, Cousins AB, Kellogg EA, Brutnell TP** (2016) The draft genome of the C3 panicoid grass species *Dichanthelium oligosanthes*. *Genome Biol* **17**: 223
- Sun G, Wase N, Shu S, Jenkins J, Zhou B, Torres-Rodríguez JV, Chen C, Sandor L, Plott C, Yoshinga Y, et al** (2022) Genome of *Paspalum vaginatum* and the role of trehalose mediated autophagy in increasing maize biomass. *Nat Commun* **13**: 7731
- Takeda T, Amano K, Ohto M, Nakamura K, Sato S, Kato T, Tabata S, Ueguchi C** (2006) RNA Interference of the Arabidopsis Putative Transcription Factor TCP16 Gene Results in Abortion of Early Pollen Development. *Plant Mol Biol* **61**: 165–177
- Tanaka H, Hirakawa H, Kosugi S, Nakayama S, Ono A, Watanabe A, Hashiguchi M, Gondo T, Ishigaki G, Muguerza M, et al** (2016) Sequencing and comparative analyses of the genomes of zoysiagrasses. *DNA Res* **23**: 171–180
- Tebele SM, Marks RA, Farrant JM** (2021) Two Decades of Desiccation Biology: A Systematic Review of the Best Studied Angiosperm Resurrection Plants. *Plants* **10**: 2784
- Tsai KJ, Lu M-YJ, Yang K-J, Li M, Teng Y, Chen S, Ku MSB, Li W-H** (2016) Assembling the *Setaria italica* L. Beauv. genome into nine chromosomes and insights into regions affecting growth and drought tolerance. *Sci Rep* **6**: 35076
- VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, Chaluvadi SR, Han G, Bryant D, Edger PP, et al** (2020) Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat Commun* **11**: 884
- VanBuren R, Pardo J, Man Wai C, Evans S, Bartels D** (2019) Massive Tandem Proliferation of ELIPs Supports Convergent Evolution of Desiccation Tolerance across Land Plants.

Plant Physiol **179**: 1040–1049

- VanBuren R, Wai CM, Keilwagen J, Pardo J** (2018) A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* **2**: e00096
- VanBuren R, Wai CM, Zhang Q, Song X, Edger PP, Bryant D, Michael TP, Mockler TC, Bartels D** (2017) Seed desiccation mechanisms co-opted for vegetative desiccation in the resurrection grass *Oropetium thomaeum*. *Plant Cell Environ* **40**: 2292–2306
- Vecchia FD, El Asmar T, Calamassi R, Rascio N, Vazzana C** (1998) Morphological and ultrastructural aspects of dehydration and rehydration in leaves of *Sporobolus stapfianus*. *Plant Growth Regul* **24**: 219–228
- Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, et al** (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**: 763–768
- Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej MC, Delorean E, Thambugala D, et al** (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature* **588**: 277–283
- Wang H, Leng X, Xu X, Li C** (2020) Comprehensive Analysis of the TIFY Gene Family and Its Expression Profiles under Phytohormone Treatment and Abiotic Stresses in Roots of *Populus trichocarpa*. *Forests* **11**: 315
- Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, Zuccolo A, Song X, Kudrna D, Ammiraju JSS, et al** (2014a) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* **46**: 982–988
- Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, et al** (2014b) The *Spirodela polyrrhiza* genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nat Commun* **5**: 3311
- Wang X, Chen S, Ma X, Yssel AEJ, Chaluvadi SR, Johnson MS, Gangashetty P, Hamidou F, Sanogo MD, Zwaenepoel A, et al** (2021a) Genome sequence and genetic diversity analysis of an under-domesticated orphan crop, white fonio (*Digitaria exilis*). *GigaScience* **10**: giab013
- Wang X, Niu Y, Zheng Y** (2021b) Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. *Int J Mol Sci* **22**: 6125
- Willig J-J, Guarneri N, Steenbrugge JJM van, Jong W de, Chen J, Goverse A, Torres JLL, Sterken MG, Bakker J, Smant G** (2022) The *Arabidopsis* transcription factor TCP9 modulates root architectural plasticity, reactive oxygen species-mediated processes, and tolerance to cyst nematode infections. *Plant J* **112**: 1070–1083

- Yan J, Liu Y, Yang L, He H, Huang Y, Fang L, Scheller HV, Jiang M, Zhang A (2021a)** Cell wall  $\beta$ -1,4-galactan regulated by the BPC1/BPC2-GALS1 module aggravates salt sensitivity in *Arabidopsis thaliana*. *Mol Plant* **14**: 411–425
- Yan N, Yang T, Yu X-T, Shang L-G, Guo D-P, Zhang Y, Meng L, Qi Q-Q, Li Y-L, Du Y-M, et al (2022)** Chromosome-level genome assembly of *Zizania latifolia* provides insights into its seed shattering and phytocassane biosynthesis. *Commun Biol* **5**: 1–11
- Yan Q, Wu F, Xu P, Sun Z, Li J, Gao L, Lu L, Chen D, Muktar M, Jones C, et al (2021b)** The elephant grass (*Cenchrus purpureus*) genome provides insights into anthocyanidin accumulation and fast growth. *Mol Ecol Resour* **21**: 526–542
- Yoon Y, Seo DH, Shin H, Kim HJ, Kim CM, Jang G (2020)** The Role of Stress-Responsive Transcription Factors in Modulating Abiotic Stress Tolerance in Plants. *Agronomy* **10**: 788
- Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, et al (2011)** The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520–524
- Zhang J, Wu F, Yan Q, John UP, Cao M, Xu P, Zhang Z, Ma T, Zong X, Li J, et al (2021)** The genome of *Cleistogenes songorica* provides a blueprint for functional dissection of dimorphic flower differentiation and drought adaptability. *Plant Biotechnol J* **19**: 532–547
- Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, et al (2018)** Allele-defined genome of the autoployploid sugarcane *Saccharum spontaneum* L. *Nat Genet* **50**: 1565–1573
- Zhang X, Ran W, Zhang J, Ye M, Lin S, Li X, Sultana R, Sun X (2020)** Genome-Wide Identification of the Tify Gene Family and Their Expression Profiles in Response to Biotic and Abiotic Stresses in Tea Plants (*Camellia sinensis*). *Int J Mol Sci* **21**: 8316
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al (2008)** Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137
- Zhao Z, Meng G, Zamin I, Wei T, Ma D, An L, Yue X (2023)** Genome-Wide Identification and Functional Analysis of the TIFY Family Genes in Response to Abiotic Stresses and Hormone Treatments in Tartary Buckwheat (*Fagopyrum tataricum*). *Int J Mol Sci* **24**: 10916
- Zhu T, Wang L, Rimbart H, Rodriguez JC, Deal KR, De Oliveira R, Choulet F, Keeble-Gagnère G, Tibbits J, Rogers J, et al (2021)** Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. *Plant J* **107**: 303–314

**Zimin AV, Puiu D, Hall R, Kingan S, Clavijo BJ, Salzberg SL** (2017) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *GigaScience* **6**: gix097

**Zou C, Massonnet M, Minio A, Patel S, Llaca V, Karn A, Gouker F, Cadle-Davidson L, Reisch B, Fennell A, et al** (2021) Multiple independent recombinations led to hermaphroditism in grapevine. *Proc Natl Acad Sci* **118**: e2023548118

APPENDIX A: FIGURES AND TABLES

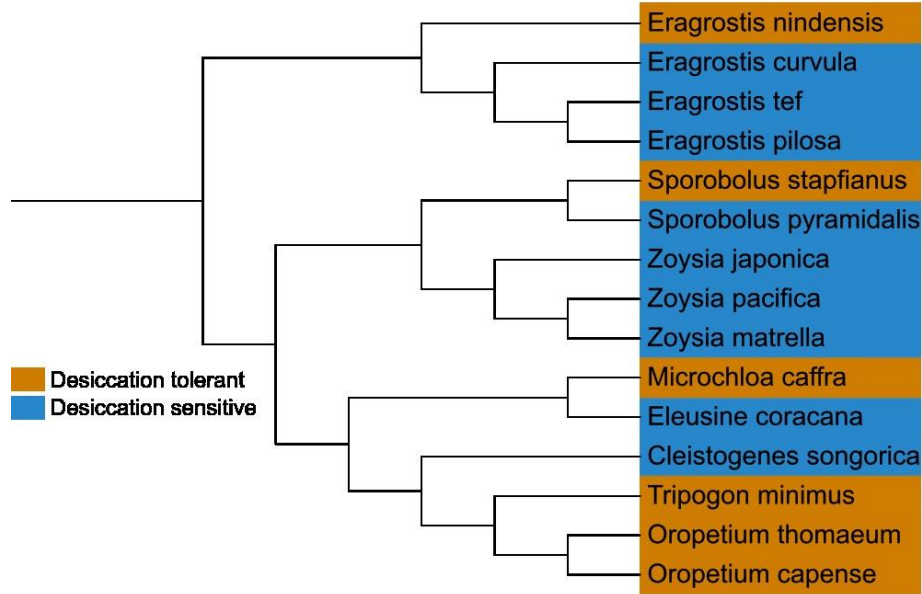


Figure A4.1: Species tree (generated using OrthoFinder) of the chloridoid grasses used in this study. Desiccation tolerant (DT) species are highlighted in orange, desiccation sensitive (DS) in blue.



Table A4.1: Selected expanded orthogroups in DT compared to DS chloridoids, with the functional names of their *Arabidopsis* or rice ortholog(s). These “expanded orthogroups of particular interest” were selected because they were less likely to be phylogenetic artifacts, or because they had differential expression in all six DT species studied. Orthogroups with the lowest p-values are at the top of the table.

<b>Orthogroup</b>	<b>Arabidopsis or rice ortholog(s)</b>	<b>Mean number of copies in DT species</b>	<b>Mean number of copies in DS species</b>	<b>Expansion p-value</b>
OG0000195	Myb/SANT-like DNA-binding domain proteins	61	11	2.25E-06
OG0000413	Early light-induced proteins	27	15	0.00054
OG0000047	Subtilisin proteases, etc.	33	25	0.00136
OG0000067	Pathogenesis-related thaumatin superfamily proteins	27	23	0.00178
OG0000164	Expansin precursors (putative, expressed)	19	15	0.00213
OG0000055	Protein kinases	28	25	0.00266
OG0000033	Protein kinases	25	24	0.00484
OG0000613	Low temperature and salt responsive proteins	14	12	0.00513
OG0000050	Resistance proteins	23	15	0.00539
OG0000355	Cyclin T1	11	8	0.00587
OG0002131	VQ motif-containing proteins	8	5	0.00781
OG0000490	alpha/beta-Hydrolases superfamily proteins	11	8	0.0081
OG0000307	Chitinase family proteins	14	12	0.00818
OG0003146	ASYNAPTIC4	6	3	0.00836
OG0000240	CCR-associated factor deadenylases	16	16	0.00844

Table A4.1 (cont'd)

OG0000273	response regulators involved in cytokinin- mediated signaling	22	21	0.00916
OG0000418	ARIADNE/ARI/ATAR I RING/U-box superfamily proteins	9	7	0.00967

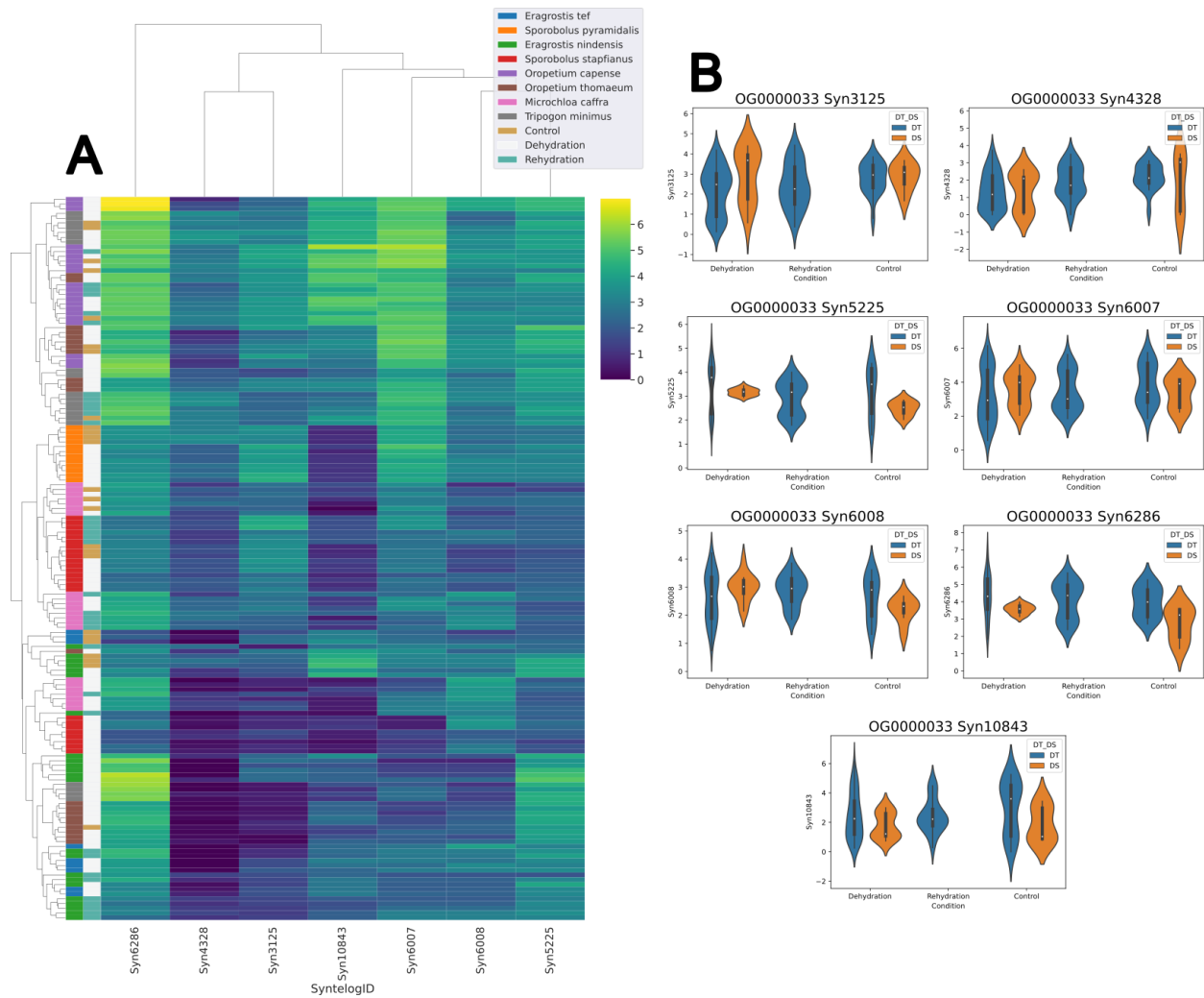


Figure A4.2: Expression of conserved syntelogs in OG0000033, protein kinases, plotted on a heatmap (A) and as individual violin plots for each syntelog (B). Notably, syntelog 6286 was, on average, more highly expressed under both desiccation and control in DT than DS species.

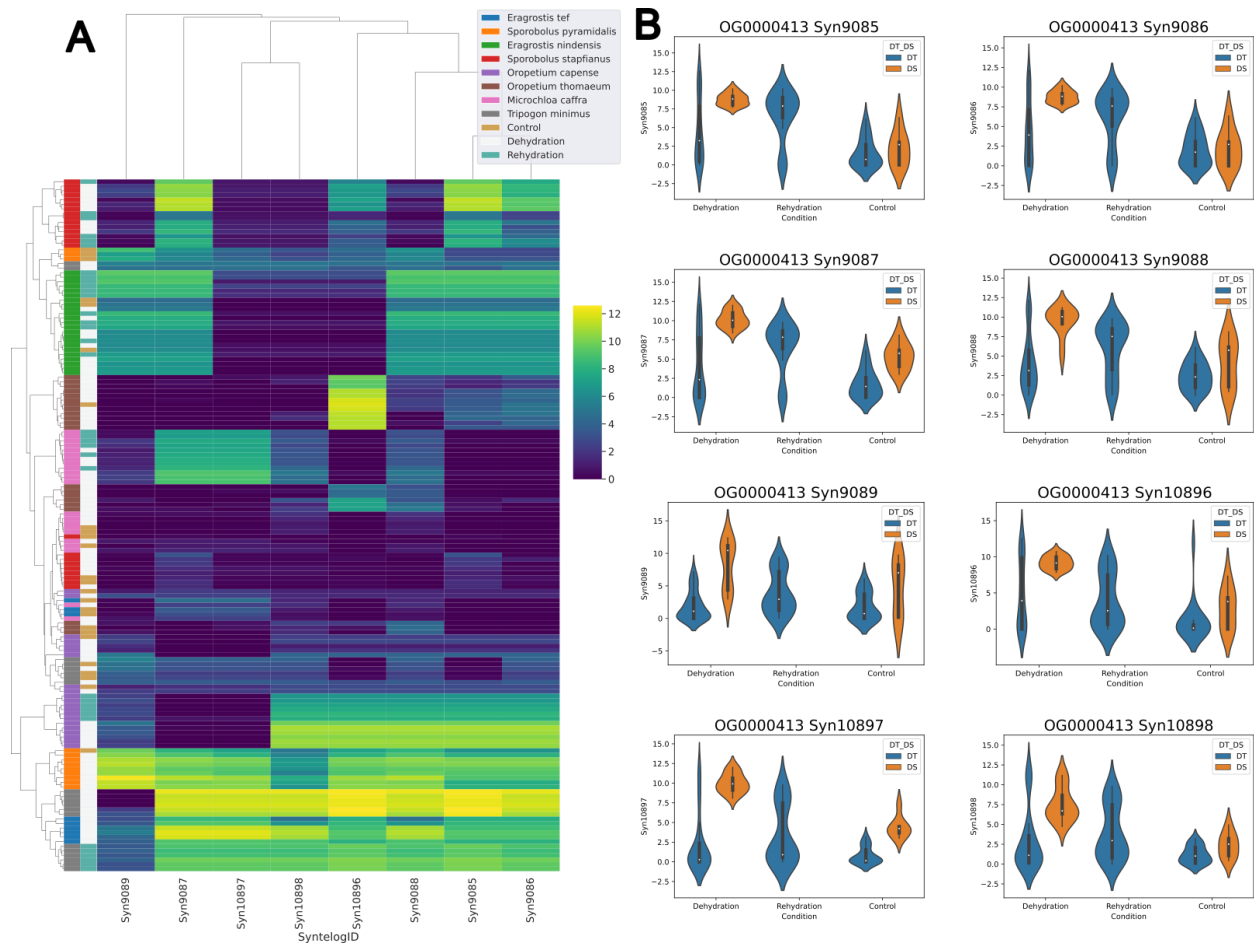


Figure A4.3: Expression of conserved syntenic orthologs in OG0000413, early light-induced proteins (ELIPs) as heatmap (A) and violin plots of individual syntelogs' expression under different conditions in DT and DS species (B).

Table A4.2: Motifs enriched in promoters of DEGs across the six DT and/or two DS chloridoid species, along with their associated transcription factors (motif names). Of particular interest are the motifs conserved across DT species but not conserved across all species regardless of tolerance.

<b>Motif name from HOMER</b>	<b>Consensus sequence</b>	<b>Set(s) of DEGs</b>	<b>Conserved across DT species?</b>	<b>Conserved across DT &amp; DS species?</b>
ZIM(C2C2gata)	ATCSRACGGTYRA GA	Up & down	Yes	No
ABF2	KGMCACGTGDCM HHH	Up	Yes	Yes
FRS9	RGAGAGAGAAAG	Up & down	Yes	Yes
GBF5	WKNWSACGTGGC AWN	Up	Yes	Yes
GBF6	WWTGMCACGTCA BCW	Up	Yes	Yes
bZIP16	TGCCACGTGD	Up	Yes	Yes
bZIP28	TGCCACGTSABH	Up	Yes	Yes
bZIP48	DDWWKVTSACGTG GC	Up	Yes	Yes
bZIP53	NDNHSACGTGKMN NN	Up	Yes	Yes
bZIP68	WGCCACGTGK	Up	Yes	Yes
At1g72010(TC P)	GGDCCCAC	Down	Yes	Yes
At2g45680(TC P)	GTGGGNCCCACND ND	Down	Yes	No
At5g08330(TC P)	GGRCCCAC	Down	Yes	Yes
BPC1	GARGAGAGAGAA	Down	Yes	No
BPC6	YTYTCTCTCTCTCT A	Down	Yes	No
PCF	NNWWWTGGGCT DDN	Down	Yes	Yes
TCP16	GTGGDCCYNNNNN NN	Down	Yes	No

Table A4.2 (cont'd)

ZML1	ATCWYRACCGTTS RW	Down	Yes	No
ZNF692	GTGGGCCCCA	Down	Yes	Yes







	ZIM(C2C2gata)/col-ZIM-DAP-Seq(GSE60143)/Homer
	At2g45680(TCP)/colamp-At2g45680-DAP-Seq(GSE60143)/Homer
	BPC1(BBRBPC)/colamp-BPC1-DAP-Seq(GSE60143)/Homer
	BPC6(BBRBPC)/col-BPC6-DAP-Seq(GSE60143)/Homer
	TCP16(TCP)/colamp-TCP16-DAP-Seq(GSE60143)/Homer
	ZML1(C2C2gata)/colamp-ZML1-DAP-Seq(GSE60143)/Homer

Figure A4.4: Motifs enriched in promoters of DE genes across DT species, but not in DS species. See Table A4.2 for further information about regulatory direction of genes with these motifs enriched in their promoters.

APPENDIX B: SUPPLEMENTAL FIGURES AND TABLES

Table B4.1: Genomes used for OrthoFinder analysis, with taxonomic information and references.

Family	Subfamily	Tribe	Species	Genome version	Reference
Amaryllidaceae	NA	NA	<i>Allium cepa</i>		(Finkers et al., 2021)
Araceae	NA	NA	<i>Spirodela polyrhiza</i>		(Wang et al., 2014b)
Arecaceae	NA	NA	<i>Phoenix dactylifera</i>		(Al-Mssallem et al., 2013)
Asparagaceae	NA	NA	<i>Asparagus officinalis</i>		(Harkess et al., 2017)
Brassicaceae	NA	NA	<i>Arabidopsis thaliana</i>	Araport11	(Cheng et al., 2017)
Bromeliaceae	NA	NA	<i>Ananas comosus</i>		(Ming et al., 2015)
Cymodoceaceae	NA	NA	<i>Amphibolis antarctica</i>		(Bayer et al., 2022)
Cyperaceae	NA	NA	<i>Rhynchospora tenuis</i>		(Hofstatter et al., 2022)
Dioscoreaceae	NA	NA	<i>Dioscorea alata</i>		(Bredeson et al., 2022)
Fabaceae	NA	NA	<i>Medicago truncatula</i>		(Young et al., 2011)
Iridaceae	NA	NA	<i>Iris pallida</i>		(Bruccoleri et al., 2023)
Juncaceae	NA	NA	<i>Juncus inflexus</i>		(Planta et al., 2022)
Musaceae	NA	NA	<i>Musa acuminata</i>		(D'Hont et al., 2012)
Orchidaceae	NA	NA	<i>Dendrobium huoshanense</i>		(Han et al., 2020)
Poaceae	Anomochloideae	Streptochaeteae	<i>Streptochaeta angustifolia</i>		(Seetharam et al., 2021)



Table B4.1 (cont'd)

Poaceae	Bambusoidea e	Arundinarie ae	<i>Phyllostachys edulis</i>		(Ramakrishnan et al., 2020)
Poaceae	Bambusoidea e	Bambuseae	<i>Bonia amplexicaulis</i>		(Guo et al., 2019)
Poaceae	Bambusoidea e	Bambuseae	<i>Guadua angustifolia</i>		(Guo et al., 2019)
Poaceae	Bambusoidea e	Olyreae	<i>Olyra latifolia</i>		(Guo et al., 2019)
Poaceae	Bambusoidea e	Olyreae	<i>Raddia distichophylla</i>		(Li et al., 2021c)
Poaceae	Bambusoidea e	Olyreae	<i>Raddia guianensis</i>		(Guo et al., 2019)
Poaceae	Chloridoideae	Cynodonteae	<i>Cleistogenes songorica</i>		(Zhang et al., 2021)
Poaceae	Chloridoideae	Cynodonteae	<i>Eleusine coracana</i>		(Hittalmani et al., 2017)
Poaceae	Chloridoideae	Cynodonteae	<i>Microchloa caffra</i>		(Marks et al., 2024)
Poaceae	Chloridoideae	Cynodonteae	<i>Oropetium capense</i>		(Marks et al., 2024)
Poaceae	Chloridoideae	Cynodonteae	<i>Oropetium thomaeum</i>		(VanBuren et al., 2018)
Poaceae	Chloridoideae	Cynodonteae	<i>Tripogon minimus</i>		(Marks et al., 2024)
Poaceae	Chloridoideae	Eragrostideae	<i>Eragrostis curvula</i>		(Carballo et al., 2019)
Poaceae	Chloridoideae	Eragrostideae	<i>Eragrostis nindensis</i>	V3	Chapter 2 of this dissertation
Poaceae	Chloridoideae	Eragrostideae	<i>Eragrostis pilosa</i>		
Poaceae	Chloridoideae	Eragrostideae	<i>Eragrostis tef</i>	V3	(VanBuren et al., 2020)
Poaceae	Chloridoideae	Zoysieae	<i>Sporobolus pyramidalis</i>		(Chávez Montes et al., 2022)

Table B4.1 (cont'd)

Poaceae	Chloridoideae	Zoysieae	<i>Sporobolus stapfianus</i>		(Chávez Montes et al., 2022)
Poaceae	Chloridoideae	Zoysieae	<i>Zoysia japonica</i>		(Tanaka et al., 2016)
Poaceae	Chloridoideae	Zoysieae	<i>Zoysia matrella</i>		(Tanaka et al., 2016)
Poaceae	Chloridoideae	Zoysieae	<i>Zoysia pacifica</i>		(Tanaka et al., 2016)
Poaceae	Oryzoideae	Oryzeae	<i>Leersia perrieri</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza barthii</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza brachyantha</i>		(Chen et al., 2013)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza glaberrima</i>		(Wang et al., 2014a)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza glumipatula</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza longistaminata</i>		(Reuscher et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza meridionalis</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza nivara</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza punctata</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza rufipogon</i>		(Stein et al., 2018)
Poaceae	Oryzoideae	Oryzeae	<i>Oryza sativa</i>	V7	(Ouyang et al., 2007)
Poaceae	Oryzoideae	Oryzeae	<i>Zizania latifolia</i>		(Yan et al., 2022a)
Poaceae	Oryzoideae	Oryzeae	<i>Zizania palustris</i>		(Haas et al., 2021)

Table B4.1 (cont'd)

Poaceae	Panicoideae	Andropogoneae	<i>Miscanthus lutarioriparius</i>		(Miao et al., 2021)
Poaceae	Panicoideae	Andropogoneae	<i>Miscanthus sinensis</i>		(Mitros et al., 2020)
Poaceae	Panicoideae	Andropogoneae	<i>Saccharum spontaneum</i>		(Zhang et al., 2018)
Poaceae	Panicoideae	Andropogoneae	<i>Sorghum bicolor</i>		(McCormick et al., 2018)
Poaceae	Panicoideae	Andropogoneae	<i>Zea mays</i>	B73 V5	(Hufford et al., 2021)
Poaceae	Panicoideae	Paniceae	<i>Cenchrus purpureus</i>		(Yan et al., 2021b)
Poaceae	Panicoideae	Paniceae	<i>Dichanthelium oligosanthes</i>		(Studer et al., 2016)
Poaceae	Panicoideae	Paniceae	<i>Digitaria exilis</i>		(Wang et al., 2021a)
Poaceae	Panicoideae	Paniceae	<i>Echinochla crus-galli</i>		(Guo et al., 2017)
Poaceae	Panicoideae	Paniceae	<i>Panicum hallii</i>		(Lovell et al., 2018)
Poaceae	Panicoideae	Paniceae	<i>Panicum virgatum</i>		(Lovell et al., 2021)
Poaceae	Panicoideae	Paniceae	<i>Setaria italica</i>		(Tsai et al., 2016)
Poaceae	Panicoideae	Paniceae	<i>Setaria viridis</i>		(Mamidi et al., 2020)
Poaceae	Panicoideae	Paspaleae	<i>Paspalum vaginatum</i>		(Sun et al., 2022)
Poaceae	Pharoideae	Phareae	<i>Pharus latifolius</i>		(Ma et al., 2021)
Poaceae	Pooideae	Brachypodiaceae	<i>Brachypodium distachyon</i>		(Vogel et al., 2010)
Poaceae	Pooideae	Brachypodiaceae	<i>Brachypodium hybridum</i>		(Gordon et al., 2020)

Table B4.1 (cont'd)

Poaceae	Pooideae	Brachypodi eae	<i>Brachypodium stacei</i>		(Gordon et al., 2020)
Poaceae	Pooideae	Poeae	<i>Avena sativa</i>		(Kamal et al., 2022)
Poaceae	Pooideae	Poeae	<i>Lolium multiflorum</i>		(Knorst et al., 2019)
Poaceae	Pooideae	Poeae	<i>Lolium perenne</i>		(Nagy et al., 2022)
Poaceae	Pooideae	Poeae	<i>Lolium rigidum</i>		(Paril et al., 2022)
Poaceae	Pooideae	Triticeae	<i>Aegilops tauschii</i>		(Luo et al., 2017)
Poaceae	Pooideae	Triticeae	<i>Hordeum marinum</i>		(Kuang et al., 2022)
Poaceae	Pooideae	Triticeae	<i>Hordeum vulgare</i>	Morex V3	(Beier et al., 2017)
Poaceae	Pooideae	Triticeae	<i>Secale cereale</i>		(Li et al., 2021a)
Poaceae	Pooideae	Triticeae	<i>Thinopyrum intermedium</i>		
Poaceae	Pooideae	Triticeae	<i>Triticum aestivum</i>	IWGSC V2	(Zimin et al., 2017; Zhu et al., 2021)
Poaceae	Pooideae	Triticeae	<i>Triticum dicoccoides</i>		(Akpınar et al., 2018)
Poaceae	Pooideae	Triticeae	<i>Triticum spelta</i>		(Walkowiak et al., 2020)
Poaceae	Pooideae	Triticeae	<i>Triticum turgidum</i>		(Maccaferri et al., 2019)
Poaceae	Pooideae	Triticeae	<i>Triticum urartu</i>		(Ling et al., 2018)
Posidoniaceae	NA	NA	<i>Posidonia australis</i>		(Bayer et al., 2022)
Solanaceae	NA	NA	<i>Solanum lycopersicum</i>	V4	(Mueller et al., 2009)

Table B4.1 (cont'd)

Velloziaceae	NA	NA	<i>Xerophyta schlechteri</i>		(Costa et al., 2017)
Vitaceae	NA	NA	<i>Vitis vinifera</i>		(Zou et al., 2021)
Zingiberaceae	NA	NA	<i>Zingiber officinale</i>		(Li et al., 2021b)
Zosteraceae	NA	NA	<i>Zostera marina</i>		(Olsen et al., 2016)

Table B4.2: Description of RNA-seq samples used for each species. In most cases, we used previously calculated differentially expressed genes (DEGs), but for *E. nindensis*, we re-processed the samples with the V3 genome (see chapter 2 of this dissertation). Only leaf samples were used. gH<sub>2</sub>O: grams of water; gDW: grams dry weight.

<b>Species</b>	<b>DT or DS?</b>	<b>Description of samples</b>	<b>Reference</b>
<i>Eragrostis tef</i>	DS	Well-watered and drought (128 and 152 hours)	(Pardo et al., 2020)
<i>Sporobolus pyramidalis</i>	DS	Well-watered (3 gH <sub>2</sub> O/gDW) and drought (2 and 1.5 gH <sub>2</sub> O/gDW)	(Chávez Montes et al., 2022)
<i>Eragrostis nindensis</i>	DT	Well-watered, desiccation (56, 104, and 228 hours), and rehydration (12, 24, and 48 hours)	(Pardo et al., 2020)
<i>Microchloa caffra</i>	DT	Well-watered, desiccation (120, 216, 264, and 432 hours), and rehydration (24 and 48 hours)	(Marks et al., 2024)
<i>Oropetium capense</i>	DT	Well-watered, desiccation (144, 240, 336, and 480 hours), and rehydration (24 and 48 hours)	(Marks et al., 2024)
<i>Oropetium thomaeum</i>	DT	Well-watered, desiccation (7, 14, 21, and 30 days), and rehydration (24 and 48 hours)	(VanBuren et al., 2017)
<i>Sporobolus stapfianus</i>	DT	Well-watered (3 gH <sub>2</sub> O/gDW), desiccation (2, 1.5, 1, 0.75, and 0.5 gH <sub>2</sub> O/gDW), and rehydration (12 and 24 hours)	(Chávez Montes et al., 2022)
<i>Tripogon minimus</i>	DT	Well-watered, desiccation (144, 240, 336, and 480 hours), and rehydration (24 and 48 hours)	(Marks et al., 2024)

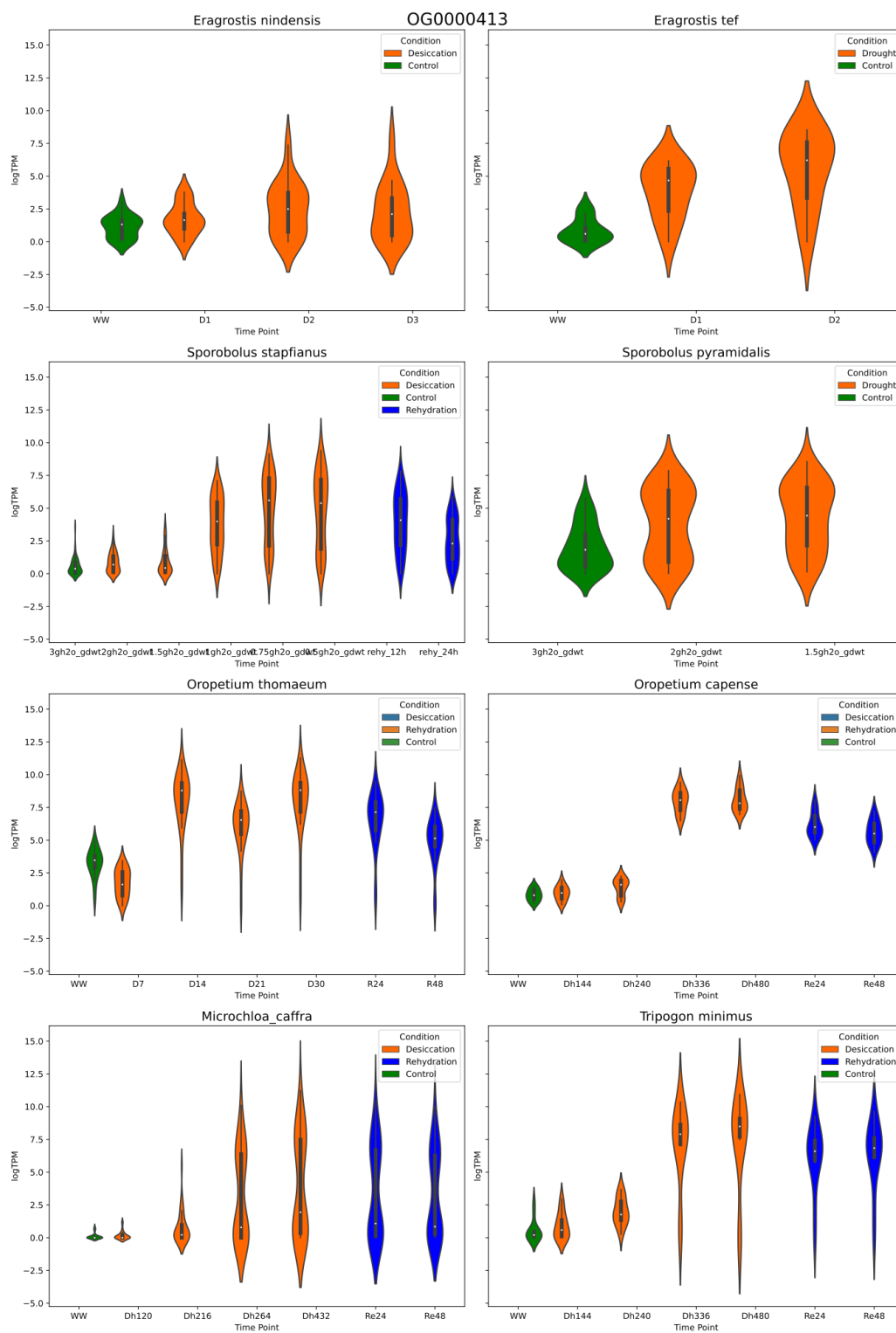


Figure B4.1: Violin plots of the expression of OG0000413, early light-induced proteins (ELIPs), in each of the six DT and two DS species examined. In general, regardless of the species' desiccation tolerance, ELIP expression increases during dehydration.

Table B4.3: Numbers of predicted open chromatin (ATAC-seq) peaks from Predmoter for each genome, along with ploidy of each species. Additionally for *O. thomaeum*, there were 26,321 real ATAC-seq peaks identified by (St. Aubin et al., 2022), the vast majority of which overlapped with peaks from Predmoter.

<b>Species</b>	<b>Ploidy</b>	<b>Number of peaks predicted</b>
<i>Oropetium thomaeum</i>	Diploid	143,049
<i>Oropetium capense</i>	Diploid	137,853
<i>Eragrostis nindensis</i>	Tetraploid	331,599
<i>Eragrostis tef</i>	Tetraploid	234,137
<i>Sporobolus stapfianus</i>	Tetraploid	420,639
<i>Sporobolus pyramidalis</i>	Hexaploid	374,474
<i>Microchloa caffra</i>	Hexaploid	392,973
<i>Tripogon minimus</i>	Diploid	115,084



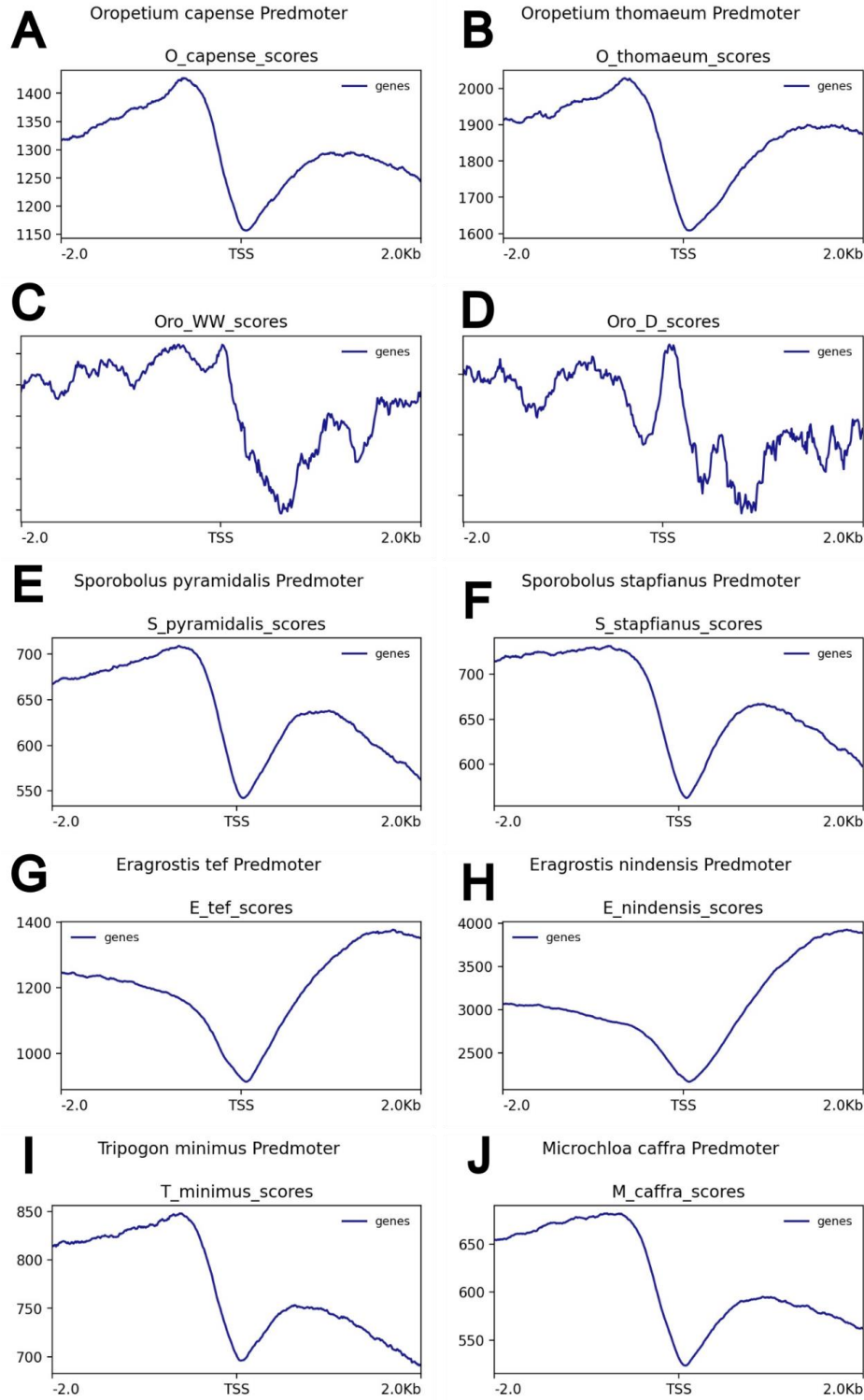


Figure B4.2: deepTools plotProfile results showing distributions of predicted (A-B and E-J) and real (C-D) ATAC-seq peaks relative to in the transcription start sites (TSS) of genes in the various species. In almost all cases, the most peaks are present before the TSS, as expected.

Table B4.4: Numbers of differentially expressed genes (DEGs) in each species analyzed. For most species, these DEGs were calculated in previous papers (see Table B4.2 for citations); for *E. nindensis*, they were calculated here after raw RNA-seq data was re-analyzed against the *E. nindensis* V3 genome.

<b>Species</b>	<b>Number of total DEGs</b>	<b>Number of upregulated genes</b>	<b>Number of downregulated genes</b>
<i>Oropetium thomaeum</i>	16,123	8,157	9,229
<i>Oropetium capense</i>	15,942	8,178	9,868
<i>Eragrostis nindensis</i>	12,332	4,951	7,396
<i>Eragrostis tef</i>	9,520	5,538	3,982
<i>Sporobolus stapfianus</i>	43,794	20,126	27,257
<i>Sporobolus pyramidalis</i>	27,567	13,313	14,265
<i>Microchloa caffra</i>	31,564	13,180	19,737
<i>Tripogon minimus</i>	14,678	6,730	8,487

## **Future Directions**

The work put forward in this dissertation is hopefully an important contribution to the broader field of abiotic stress genomics of grasses, particularly in terms of gene expression regulation. In studying both core stress response in maize and conserved genomic mechanisms of desiccation tolerance in chloridoid grasses, we have identified several important regulators of various transcription factor families, namely MYB, BBR/BPC, TCP, TIFY, ERF, bZIP, NAC, C2C2-CO-like, and HSF. All these families have previously been identified as stress-responsive. We identified MYBs, BBR/BPCs, TCPs, and TIFs as putative regulators of desiccation tolerance (Chapter 4), and only MYBs have been related to desiccation tolerance before, albeit in non-grass desiccation tolerant plants (Ma et al., 2015; Zhang et al., 2024). Thus, we have identified several novel regulators of desiccation tolerance in chloridoid grasses, which may be important in other desiccation tolerant organisms as well. We further identified ERFs, bZIPs, NACs, HSFs, and C2C2-CO-like transcription factors as enriched in core stress genes in maize (Chapter 2). While these gene families have all been previously linked to abiotic stress response, this is the first study to our knowledge that has explicitly identified these transcription factors as key regulators of all six stressors studied in Chapter 2. Thus, we have contributed significantly to the knowledge base on regulation of abiotic stress-related gene expression.

In Chapter 1, core abiotic stress response genes were identified from public RNA-seq data using two methods, random forest classification and set operations on differentially expressed genes. We concluded that core genes in certain transcription factor families, namely AP2/ERF-ERF, bZIP, NAC, C2C2-CO-like, and HSF, likely regulate both other core genes and certain sets of stress-specific genes. Before this work is peer-reviewed and published, we plan to add further analyses to provide support for this hypothesis. For instance, we will identify the genes whose promoters contain motifs for each of the core transcription factors of interest using FIMO in the MEME suite (Bailey et al., 2015), and use Fisher's exact test to find whether there are more core genes or stress-specific genes (each stressor, testing up- and downregulated genes separately) regulated by each transcription factor than expected by chance. Similarly, building a protein-protein interaction network with STRING (Szklarczyk et al., 2023) for the core genes will enable us to identify any physical interactions between the core transcription factors which may be involved in their regulatory activities. In similar future studies, it would be ideal to

construct a gene regulatory network to better understand the regulation of core and peripheral stress genes.

There are multiple potentially interesting ways to expand upon the work presented in Chapter 1. For instance, the existing meta-analysis could be expanded to cover more species, with comparison of species-specific and cross-species core genes. It would be interesting to compare core genes from generally stress-sensitive major crop and model species, for instance, with those from more stress-tolerant species like sorghum, or even extremophiles that are highly tolerant to certain stress conditions. However, most species have not been as thoroughly studied as maize, necessitating the generation of data via a controlled environment core stress experiment. Although this would be valuable for a wild extremophile, such as a desiccation tolerant plant, for which little is known about its performance under stressors other than desiccation, a core stress experiment could also be conducted for maize. It would be useful to compare the results of a maize core stress experiment with the results of the meta-analysis conducted here, to find out how much the core gene sets overlap between them.

As mentioned in Chapter 2, although the *Eragrostis nindensis* genome is now greatly improved in contiguity and we have determined the likely polyploid origin of the species, the genome is still not chromosome scale. Generation of new Hi-C data from the same individual plant (still maintained in the VanBuren lab) and use of this data for scaffolding could assist in resolving the genome to chromosome scale.

In Chapter 3, we concluded that both gene family expansion and changes in gene expression regulation have contributed to the evolution of desiccation tolerance in the Chloridoideae. The statistical method used to test orthogroup expansion in Chapter 3 was selected due to its ability to correct for the ploidy level of different species (see Chapter 3, Methods). However, this method does not account for phylogeny. Using a program such as CAFE5 (Mendes et al., 2021) would identify not only expanded orthogroups, but the nodes of the phylogenetic tree at which they were expanded, providing important information about timing of expansion in various lineages. Using this method will also enable us to determine whether orthogroup expansions were convergent in DT species or whether the expansion originally occurred in the last common ancestor of current chloridoids, and was subsequently lost in DS species. We plan to implement this method to re-test orthogroup expansion before publishing this work.

In Chapter 3, we generated orthogroups for 67 grass species, as well as 16 monocots and 4 eudicots as outgroups. Thus, although in Chapter 3 we focused only on the 15 Chloridoideae species included, this dataset can be used for similar comparisons with other groups of grasses. For instance, we could find expanded orthogroups in the desiccation sensitive species of the Chloridoideae compared to non-chloridoid grasses, and compare these with the expanded orthogroups found in desiccation tolerant chloridoids; this analysis could provide insight into the resilience of chloridoids compared to the “super-resilience” of desiccation tolerant chloridoids. Similar orthogroup expansion analyses could be conducted for salt-tolerant grasses or stress-tolerant grasses in general; this could be another approach to finding cross-species core mechanisms of stress tolerance.

In sum, while we hope the work presented in this dissertation contributes to the field of grass abiotic stress biology, there is room for further research. We hope that our findings, including candidate stress tolerance genes identified here as well as broader patterns of core stress response, will be built upon in future by other researchers to better understand plant abiotic stress response.

## REFERENCES

- Bailey TL, Johnson J, Grant CE, Noble WS** (2015) The MEME Suite. *Nucleic Acids Res* **43**: W39–W49
- Ma C, Wang H, Macnish AJ, Estrada-Melo AC, Lin J, Chang Y, Reid MS, Jiang C-Z** (2015) Transcriptomic analysis reveals numerous diverse protein kinases and transcription factors involved in desiccation tolerance in the resurrection plant *Myrothamnus flabellifolia*. *Hortic Res*. doi: 10.1038/hortres.2015.34
- Mendes FK, Vanderpool D, Fulton B, Hahn MW** (2021) CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**: 5516–5518
- Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al** (2023) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**: D638–D646
- Zhang X, Ekwealor JTB, Mishler BD, Silva AT, Yu L, Jones AK, Nelson ADL, Oliver MJ** (2024) *Syntrichia ruralis*: emerging model moss genome reveals a conserved and previously unknown regulator of desiccation in flowering plants. *New Phytol*. doi: 10.1111/nph.19620