

DISCOVERY OF DOMESTICATION, RESILIENCE, AND AGRONOMIC TRAITS IN THE  
UNDERUTILIZED CEREAL TEFF

By

McKena Lipham Wilson

A DISSERTATION

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

Plant Breeding, Genetics, and Biotechnology – Horticulture – Doctor of Philosophy

2024

## ABSTRACT

The Ethiopian grain teff (*Eragrostis tef*) is an economically and culturally important grain in the Horn of Africa, where it is most commonly grown by small scale farmers and has been domesticated to maintain consistent yields in poor conditions with low rainfall and less management. Advances in teff breeding have been slow due to its high selfing rate and lack of genetic resources. The goal of this work was to assemble a manageable panel of teff germplasm with maximum genetic diversity, leverage field phenotyping to identify marker trait associations, and share these tools and resources to advance teff breeding. Here, we describe the construction and genotyping of the Teff Association Panel (TAP), consisting of 265 cultivars and farmer varieties, as well as the wild progenitor *Eragrostis pilosa*. Using whole genome resequencing, we identified 21 million single nucleotide polymorphisms and insertions/deletions across the diversity panel, and used this panel to confirm the wild progenitor of teff, survey the genetic diversity and domestication history, and identify genetic loci underlying important agronomic traits. We grew the TAP in the field at Michigan State University in 2021 and 2022 to evaluate for lodging susceptibility, panicle architecture, plant height, days to heading, culm width, average panicle weight, average seed weight per panicle, seed color, and 12 seed mineral nutrients. The associations between agronomic and nutritional traits were evaluated to determine which traits play a critical role in lodging susceptibility and seed mineral nutrient content in teff. We detected a high correlation between panicle architecture and lodging susceptibility, and found that accessions with open panicle architectures lodged more consistently. We also confirmed the high nutritional value of teff by conducting the first large-scale analysis of teff seed mineral nutrients. The phenotypic data was harnessed to perform genome wide association for each trait and identify 50 loci significantly associated with 13 traits. The phenotypic variability and genomic resources developed from this work can be applied to rapidly improve teff agronomic efficiency.

Copyright by  
MCKENA LIPHAM WILSON  
2024

## ACKNOWLEDGEMENTS

What a whirlwind. Thank you to all who have supported me along this five year journey of learning and unlearning. This PhD has taught me so much about plant breeding and genetics, computational science, and plant physiology, but it has also taught me a tremendous amount about myself and how I want to improve over the next five years of my career.

Thank you first and foremost to Dr. Robert VanBuren. I rotated in Bob's lab as a first year, after I had decided I wanted to pursue something outside of my undergraduate field of biotechnology. He welcomed me and excitedly shared his passion for plant resilience, genomic inquiry, and his newly established lab at Michigan State University. Bob allowed me to grow as a scientist by pursuing new skills and exploring an understudied crop system. Thank you for always putting students' mental health at the forefront in your lab and for leading with humility and curiosity.

Thank you to Dr. Rose Marks for your mentorship as well as your friendship. This combination provided a unique opportunity for respectful criticisms; thank you for speaking to me frankly when needed and offering sound advice at a moment's notice.

To the rest of the VanBuren lab, past and present: Dr. Jennifer Wai, Dr. Jeremy Pardo, Dr. Anna Pardo, and Dr. Brian St. Aubin, thank you for introducing me to the VanBuren lab, and thank you for sharing your skills, your insights, and your advice freely time after time. Jenny Schuster, thank you for showing up. You helped me conquer one of the most difficult times of my PhD simply by continuing to come in and work next to me; so happy we put our brains together and moved in next door. Thank you for reminding me I can do hard things. Dr. Serena Lotreck, a bookend partner in crime, your proactive positivity is infectious, and I will always be grateful for your support. Cathy Mercado, Dr. Ian Gilman, Elliot Braun and Maddy Creach, thank you for bringing new life, ideas, and joy into the lab. So happy to have been surrounded by such amazing people. I also want to thank the team of undergraduate research assistants who helped me maintain the teff fields, collect data, and develop new protocols for teff management. I hope I taught you as much as you taught me.

Thank you to my committee. Dr. Addie Thompson, thank you for always making time for me and providing a refreshing perspective. You have helped me discover my own solution more than once, and I hope to pay this forward to someone soon. Dr. Isaacs, you are such an inspiration, thank you for always highlighting the importance of our applications in plant breeding; your

teachings have sparked my ambitions for this next stage of my career. Dr. Gomez, thank you for sharing your passion for quantitative genetics. You reignited my excitement for plant breeding at a time when the pandemic had diminished my fervor. Dr. Josephs, I learned so much from just rotating in your lab, and I'm so glad you were able to bookend my experience in graduate school. To all of you, thank you for your continued support.

I'd also like to thank Dr. Ana Maria Heilman, my NAPB Borlaug Scholar mentor. Thank you for prioritizing my professional development and broadening my network; I am truly grateful.

Thank you to all the amazing people I was able to collaborate with at Michigan State and across the globe. Thank you HOGS, Plant Biology GSO, PLB801, IMPACTS, PBGB and so many other acronyms that helped me find camaraderie among colleagues, learn to collaborate more effectively, and make meaningful changes to the grad school experience. Special thanks to Thilani Jayakody, Emily Conway, and Jared Gregorini for their friendship, support, and motivation.

Thank you to my family, all the Liphams and Wilsons. When we moved up here, we left a lot of love in Georgia, but I always felt so supported from afar. Specifically thank you to my parents for instilling in me a love for the outdoors and a strong work ethic. I didn't realize until I learned more about agriculture that I was blessed with nuggets of horticulture and plant biology throughout my childhood. Big thanks to my sister for her continual guidance and for giving me the courage to apply for graduate school in the first place.

Finally, to my husband, Blake Wilson, who truly made this possible, you are the most perfect partner. Thank you for your calm, for your encouragement, and your personal dedication to all things; you inspire me daily.

## TABLE OF CONTENTS

CHAPTER 1: LEVERAGING MILLETS FOR DEVELOPING CLIMATE RESILIENT AGRICULTURE .....	1
CHAPTER 2: A GENOMC DISCOVERY PLATFORM FOR ACCELERATING TRAIT DISCOVERY OF THE CLIMATE RESILIENT ETHIOPIAN CEREAL TEFF.....	3
REFERENCES.....	22
CHAPTER 3: GENOME WIDE ASSOCIATION OF LODGING SUSCEPTIBILITY AND PANICLE ARCHITECTURE TRAITS IN TEFF.....	25
REFERENCES.....	62
CHAPTER 4: GENETIC INSIGHTS AND EXAMINATION OF SEED NUTRIENT VARIABILITY IN THE TEFF ASSOCIATION PANEL.....	66
REFERENCES.....	89
CHAPTER 5: FUTURE DIRECTIONS.....	92

## **CHAPTER 1: LEVERAGING MILLETS FOR DEVELOPING CLIMATE RESILIENT AGRICULTURE**

The work presented in this chapter is part of the final publication:

Wilson ML, VanBuren R. Leveraging millets for developing climate resilient agriculture. Current Opinion in Biotechnology. <https://doi.org/10.1016/j.copbio.2022.102683>

Author contributions:

Both MLW and RV conceptualized and prepared the manuscript.

## **Abstract**

C4 grasses dominate natural and agricultural settings, and the widespread success of wild grasses is mostly attributable to their resilience to environmental extremes. Much of this natural stress tolerance has been lost in major cereals as a byproduct of domestication and intensive selection. Millets are an exception, and they were domesticated in semi-arid regions of Sub-Saharan Africa and Asia where selection favored tolerance and stability over yield. Here, we review the evolutionary and domestication histories of millets and the traits that enable their stress tolerance, broad adaptability, and superior nutritional qualities compared to other cereals. We discuss genome editing and advanced breeding approaches that can be used to develop nutritious, climate resilient cereals of the future. Finally, we propose that millets can play a central role in the global food system to combat food insecurity, with researchers and germplasm from the Global South at the center of these efforts.

## **Summary**

Millets were domesticated with an emphasis on stress tolerance and stability over yield. This review explores the history and traits of millets that contribute to their adaptability, nutritional content, and stress tolerance. My advisor, RV, and I conceptualized and prepared the manuscript as a literature review. Together we decided which agronomic, physiological, and nutritional traits to highlight, summarized the genomic and breeding resources that have been developed for these crops, then suggested advances to be made.



## **CHAPTER 2: A GENOMC DISCOVERY PLATFORM FOR ACCELERATING TRAIT DISCOVERY OF THE CLIMATE RESILIENT ETHIOPIAN CEREAL TEFF**

The work presented in this chapter is being prepared for publication:

Wai CM\*, Wilson ML\*, Chanyalew S, Dell'Acqua M, Thompson A, VanBuren R.

A genomic diversity platform for accelerating trait discovery of the climate resilient Ethiopian  
cereal teff. *In prep.*

Author contributions: MLW, CMW, and RV designed the experiments. MLW, CMW, and RV completed experiments and data analysis. CS, MDA, and AT provided expertise and guidance on data analysis and interpretation. MLW and RV wrote the manuscript. All authors will edit the manuscript.

## Abstract

Teff (*Eragrostis tef*) is an economically and culturally important grain crop in the Horn of Africa where thousands of locally adapted cultivars are grown primarily by small scale farmers. Teff is resilient to low rainfall, minimal inputs, and basic management practices, but lower yielding than other cereals. Here, we constructed and resequenced the Teff Association Panel (TAP) and used these 265 diverse accessions to explore the domestication and improvement history of teff. Through phylogenetic, admixture, and population differentiation analyses, we confirm that *Eragrostis pilosa* is the direct ancestor of teff, with domestication most likely occurring from a distinct population in Ethiopia's Tigray region. Comparative genomic analysis with a high quality *E. pilosa* genome revealed minimal gene loss or structural rearrangements during teff domestication, though we observed a three fold reduction in nucleotide diversity. Through genome-wide association studies, we identified two primary loci responsible for seed color, including an ortholog to CYP75B1, a cytochrome P450 enzyme involved in flavonoid biosynthesis. A missense mutation in this gene during domestication likely led to the development of white-seeded teff varieties. Together, these genetic resources can be used to accelerate agronomic trait improvement in teff.

## Introduction

Humans domesticated dozens of cereals within the Poaceae or grass family over the past 12,000 years, and cereals are a cornerstone of global food security. Cereals were domesticated in each of the historical centers of crop diversity and they originated from widely adapted and phylogenetically diverse wild grasses. Leading cereals like wheat, rice, and maize have been recurrently selected for maximum yield under optimized conditions, and the natural stress tolerance of their wild progenitors has been lost as a byproduct of selection. Millets by contrast were domesticated in semi-arid regions where selection favored stable and consistent yields under poor conditions. The domestication history of millets is rooted in indigenous practices and they are cultivated in highly diversified small-scale cropping systems that support millions of subsistence farmers. Together, these traits make millets ideal crops for developing sustainable, diversified, and climate resilient agriculture.

Teff is a primary cereal in Ethiopia where it is grown on ~24% of the total cultivated land and contributes to 21% of the yearly grain production in Ethiopia<sup>1,2</sup>. Teff provides an estimated two-thirds of the daily protein intake for most Ethiopians and it is a significant economic

commodity, as its market price is often two to three times higher than maize <sup>1,3</sup>. In Ethiopia, teff is cultivated by 6.2 million small-scale farmers and an estimated 5,000 locally adapted cultivars have been developed across the major growing regions <sup>4</sup>. This extensive local adaptation enables teff production across diverse growing conditions, including regions where major cereals may fail.

Ethiopia produces ~90-95% of the world's annual 4-5 million metric tons of teff, but the cereal has grown in popularity because of its superior nutritional profile, gluten free grain, and broad climate resilience <sup>5</sup>. Although the teff market is expected to increase, there is a history of poor research development due to a lack of funding which has restricted awareness and access to the crop <sup>5</sup>. Teff is a primarily selfing grass and outcrossing rates are as low as 1-2 percent <sup>6</sup>. Crossing teff is a technical feat still requiring manual emasculation of a 1mm flower when it opens for a concise time frame in the morning, so many researchers have implemented tilling to introduce hybridization. A CRISPR-Cas9 gene editing and transformation protocol has recently been established and can be adapted for targeted modifications in teff for breeding advancement <sup>7</sup>. Recent genomic developments, documented information on farmer preference, and sequencing of the Ethiopian diversity panel (EtDP) offer new resources and insights for molecular breeding advancement in teff <sup>8,9</sup>.

Teff has experienced exponential growth in production beyond Ethiopia, serving as a versatile alternative both as a grain and forage crop due to its superior nutritional quality, palatability, and climate resilience. As a grain crop, teff is a popular gluten free alternative with exceptional iron, calcium, and fiber content <sup>10</sup>. Although teff is highly stress tolerant, it has significantly lower yields when grown under high input conditions compared to other cereals and millets. The selection for domestication and improvement traits such as lodging tolerance, seed size, and shattering was incomplete in teff, and these issues have slowed the development of teff as a commercially viable cereal throughout the world <sup>11</sup>. To address these challenges, we have developed and sequenced a teff diversity panel to explore the domestication and improvement history of teff. These resources can serve as a foundation for improving this crucial cereal crop, optimizing its potential within Ethiopia and internationally.

## **Results**

### ***Cataloging the genetic diversity of teff***

To explore the genetic diversity of teff, we assembled a panel of 387 landraces, breeding

lines, elite cultivars, and wild *Eragrostis* germplasm from the USDA Germplasm Resources Information Network. Accessions were resequenced and aligned to the ‘Dabbi’ reference genome, with median read mapping rates of 98.8% for teff accessions, 95.0% for the putative wild progenitor *Eragrostis pilosa*, and 41.3% for all other *Eragrostis* species. Across the panel, we identified ~8.1 million single nucleotide polymorphisms, and 1.5 million insertions/deletions. The USDA GRIN germplasm was collected in the 1950s-1980s during the establishment of the Plant Genetic Resources Center of Ethiopia, and many accessions are duplicated because of insufficient passport data on origin and local variety names<sup>12</sup>. Using identity-by-descent, we found that 135 of the 363 sequenced teff lines were cryptically related and correspond to duplicated Ethiopian varieties. We retained the accession with the highest sequence coverage, resulting in a final panel of 265 unique teff lines. We refer to this refined set of germplasm as the Teff Association Panel (TAP), and we used variants within these lines and the 24 wild *Eragrostis* accessions for downstream analyses.

Teff is cultivated by millions of small-scale farmers in Ethiopia, and there is tremendous diversity found across the thousands of locally adapted and farmer-selected varieties. The Ethiopian Biodiversity Institute maintains the largest collection of teff, with approximately 6,000 accessions that encompass the global diversity of local landraces, cultivars, and lines used for teff breeding<sup>13</sup>. To assess how much of this diversity the USDA germplasm captures, we compared the genetic variation of our resequencing data to the Ethiopian Teff Diversity Panel (EtDP), 321 farmer varieties sourced from the Ethiopian Biodiversity Institute. The EtDP spans the geographical and agroecological range of teff and reflects the genetic diversity preserved within Ethiopia<sup>9</sup>. Using a common set of 7,747 SNP based markers present in both panels, we identified nine distinct subpopulations using ADMIXTURE, with each group having similar representation in the TAP and EtDP (Figure 1.1e). These subpopulations comprise between 37 to 105 teff accessions each and have high levels of admixture that reflect limited genetic stratification. Principal component analysis separates the samples by subpopulation, and accessions from both the EtDP and TAP are widely distributed across the PC1 and PC2 axes, confirming a comprehensive representation of global teff diversity in both panels (Figure 1.1a).

Accessions from the EtDP and TAP originate from all the major teff-growing regions in Ethiopia, and we observed some degree of geographic separation among subpopulation (Figure 1b). Despite this, most subpopulations exhibit broad and overlapping geographic distributions,

underscoring the extensive admixture and minimal genetic divergence within teff. This is further supported by the low fixation index ( $F_{st}$ ), a measure of genetic variation between populations, averaging 0.07 across all pairwise comparisons, indicating slight genetic separation between teff subpopulations (Figure 1.1d). The Tigray region in Northern Ethiopia is the hypothesized center of origin for teff, and we identified a geographically isolated subpopulation (7) with a significantly higher  $F_{st}$ , suggesting a distinct genetic makeup in this region (Figure 1.1d)<sup>14,15</sup>. These accessions could represent traditional varieties that have retained characteristics during early domestication as well as recurrent genetic exchanges with wild progenitors. Breeding lines curated by the Ethiopian Biodiversity Institute are predominantly found within just two subpopulations (2 and 8), indicating that a vast reservoir of genetic diversity is unutilized in teff breeding programs. This suggests an opportunity to broaden the genetic base of teff cultivars, potentially enhancing traits such as yield, resilience, and nutritional content<sup>9</sup>. This strategic expansion of the genetic pool could be pivotal for future teff improvement efforts, catering to both national and global demands for this important crop.

Teff cultivars have a range of morphological traits, and we tested if there was an association between desirable traits and genetic structure that could reflect farmer preferences. We are evaluating agronomic and nutritional traits within the panel and have found that panicle architecture, lodging tolerance, and seed color vary widely within the subpopulations, with only a few differences in trait distribution that delineate groups. Subpopulation 2 produces only white seeds, with all other subpopulations producing some combination of white, brown, and mixed color panicles.

Teff was domesticated in the Northern Ethiopian Highlands, likely from the wild grass *Eragrostis pilosa*<sup>14, 16-19</sup>. While teff and *E. pilosa* are capable of producing fertile interspecific hybrids and share similarities in numerous traits, prior studies have had too few markers or polymorphic sites to verify the origin of teff. We compared whole genome data for 12 *E. pilosa*, and other hypothesized teff progenitors including *E. heteromera*, *E. macilenta*, *E. mexicana*, and *E. papposa* to teff. Genetic analyses including admixture, phylogenetic inference, dimensionality reduction, and population differentiation ( $F_{st}$ ) provide substantial support that *E. pilosa* is the direct ancestor of teff. Maximum likelihood phylogenetic analysis of LD pruned SNPs positions *E. pilosa* sister to teff and highlights an early divergence of accessions from Tigray within the teff lineage. *E. pilosa* samples do not form a distinct group in the ADMIXTURE analysis but are

instead integrated within the genetic makeup of teff, primarily clustering within subpopulation 9, with notable admixture observed from subpopulation 7. This integration is further supported by the principal component analysis, which clusters *E. pilosa* alongside teff germplasm from both the Tigray region (subpopulation 7) and subpopulation 9, suggesting a genetic continuity between these groups. The average pairwise  $F_{st}$  between *E. pilosa* and teff subpopulations is 0.3, which is consistent with expected genetic divergence between a crop and its direct progenitor. The lowest  $F_{st}$  values were observed between *E. pilosa* and subpopulations 7 and 9. Collectively, these analyses highlight the close genetic relationship between teff and *E. pilosa*, strongly supporting the hypothesis that *E. pilosa* served as the primary ancestor of teff.

### ***Exploring the evolutionary origin of teff***

To gain a deeper insight into the origin and domestication history of teff, we constructed a *de novo* reference genome of the putative progenitor *E. pilosa* (KEW 0059857), and explored genome evolution, polyploidy, and gene level changes among these grass species. We generated 64.5 Gb of High Fidelity PacBio (HiFi) long read sequencing data and assembled the reads using Hifiasm. The resulting *E. pilosa* genome assembly is high quality, comprising 70 contigs with a total length of 560 Mb and an N50 of 14.7 Mb. Most *E. pilosa* chromosomes are assembled into 2-3 contigs and the total assembly size is similar, but ~17 Mb smaller than the ‘Dabbi’ teff reference<sup>8</sup>. Using the MAKER pipeline for *ab initio* gene prediction, we identified 69,668 gene models in *E. pilosa*, which is comparable to the 68,255 gene models annotated in ‘Dabbi’. Consistent with the near identical genome sizes, teff and *E. pilosa* have similar repetitive element composition. Repeats span 33% of the *E. pilosa* genome, with gypsy long terminal repeat retrotransposons (10.9%) and Heilitron transposons (8.5%) being the most abundant elements.

We used comparative genomics to investigate the evolutionary relationship between Eragrostis species. The A and B subgenomes of teff exhibit clear orthology to two subgenomes of *E. pilosa*, with broadly conserved macrosynteny, and no structural rearrangements (Figure 1.2). We calculated the synonymous substitution rate ( $K_s$ ) between homoeologous gene pairs within *E. pilosa* and syntenic orthologs across the two species to characterize the *E. pilosa* polyploidy event. This analysis allowed us to assign 21 and 22 contigs to A and B subgenomes of *E. pilosa*, respectively (Figure 1.2a). The median  $K_s$  values for syntenic gene pairs between the A subgenomes of *E. pilosa* and teff were 0.0036, and 0.0039 for the B subgenomes, indicating that the allotetraploidy event predates teff domestication and is shared with *E. pilosa*.

Homeologs between the *E. pilosa* A and B subgenomes have an average Ks of 0.14, and a similar Ks distribution to homeologs in teff (Figure 1.2b). Using a widely accepted mutation rate for grasses ( $1.5 \times 10^{-8}$  substitutions per nonsynonymous site per year), we estimate that this accession of *E. pilosa* and teff diverged approximately 120,000 years ago, and confirm the Eragrostis allotetraploidy event occurred ~5 million years ago <sup>8</sup>.

Consistent with their recent divergence, the teff and *E. pilosa* genomes have a high degree of collinearity with conserved gene content and order across the A and B subgenomes (Figure 1.2e). Roughly 95% of teff genes have syntenic orthologs in *E. pilosa*, with 1,702 and 1,751 teff genes having no orthologs in the corresponding *E. pilosa* A and B subgenomes (Figure 1.2c). There is no difference in fractionation or gene loss between A and B, consistent with previous observations of exceptional subgenome stability in teff <sup>8</sup>. The proportion of conserved genes between teff and its wild progenitor is considerably higher than other wild and domesticated cereals such as maize, sorghum, and rice, where half or more genes are dispensable <sup>20–22</sup>. This unusual conservation could be explained by the compact genomes, low transposable element content, and broad genome stability observed across sequenced chloridoid grasses <sup>23–26</sup>. The 3,453 teff genes that are absent from *E. pilosa* are enriched in functional roles related to core and secondary metabolism, stress responses, and metal ion transport (Figure 1.2d), and could be linked to selection during domestication.

### **Signatures of teff domestication**

Despite the vast diversity of morphological and agronomic traits, previous marker-based studies have identified limited genetic variation across teff germplasm <sup>27</sup>. Using invariant sites of reads mapped the ‘Dabbi’ reference genome, we calculated the nucleotide diversity ( $\pi$ ) within the TAP and *E. pilosa*. The genome wide nucleotide diversity of all teff accessions is  $9.7 \times 10^{-4}$  which represents a three fold reduction in diversity compared to *E. pilosa* ( $\pi = 2.9 \times 10^{-3}$ ). Nucleotide diversity is ~15% higher in the A subgenome compared to B in both teff (vs) and *E. pilosa*, which is a similar pattern to other allopolyploids including wheat <sup>28</sup>, and barnyard millet <sup>29</sup>.

To identify potential selective sweeps throughout teff domestication, we evaluated the diversity of linked regions via cross-population composite likelihood (XP-CLR) in pilosa and teff (Figure 1.3) <sup>30</sup>. A total of 222 regions were detected with an average size of 47Kb. The genomic regions contained an average of seven genes and 17% CDS. These regions together

spanned 0.27% of the CDS of the entire genome (assuming genome size of 622 Mb). A higher number of putative sweeps were detected on chromosome 10A (23), 1A (20), and 7B (19), and we detected potential collinearity within our results in corresponding subgenomes, 10B, 1B, and 7A within 500kb. Genes within the putative regions were significantly enriched in GO terms involved in pollen recognition, fungal and bacterial defense, ion transport, protein phosphorylation, oxidative stress response, and transcription.

### ***Seed color genome wide association***

To evaluate the utility of our panel for understanding the genetic basis of domestication traits in teff, we performed a genome-wide association study (GWAS) using the TAP for seed color, a key characteristic of teff. White-seeded teff is favored by growers due to its higher market value in Ethiopia; however, brown-seeded teff is noted for its superior nutritional content. *E. pilosa* produce seeds with varying shades of brown, and no white seeded *E. pilosa* lines have been observed. This suggests that brown coloration is likely the ancestral state of seed color in teff. The prevalence of white seeded teff varieties likely stems from recent selective breeding practices, influenced by consumer preferences for white seeds over brown.

Many teff varieties are grown as either mixed genotypes with different seed colors, or in rare cases, some varieties produce brown and white seeds within the same panicle. Within the TAP, 185 accessions produce brown or white seed consistently and these were used for GWA. Five loci were significantly associated with teff seed color, but two loci explain 70% of the phenotypic diversity on Chromosome 4B (13,740,406 bp; 30.6%) and Chromosome 9B (1,536,847 bp; 39.6%) (Figure 1.4). Two alleles (C and A on Chromosome 9B and 4B, respectively) are found in 79% of the brown seeded varieties, and C and T alleles at these loci are in 90% of white varieties (Figure 1.4d). Given the linkage disequilibrium (LD) decay rates of 0.1 and 0.2 at 200 Kb and 68.5 Kb respectively (Figure 1.1d), we focused on candidate genes within a 100 Kb radius. Et\_4B\_037025 is 24 Kb upstream of the loci on Chromosome 4B and is homologous to the Arabidopsis CYP75B1 gene, a flavonoid 3'-monooxygenase also known as transparent testa 7 (tt7). Arabidopsis tt7 mutants exhibit a yellow seed coat due to excessive kaempferol, compared to the dark brown seed coat observed in wild type<sup>31,32</sup>. Seed color co-segregates with a 1 bp deletion in the coding region of Et\_4B\_037025, where 87% of white accessions have the alternative allele and 87% of brown seeds have the reference allele. Over 90% of genes in the A and B subgenomes are maintained as syntenic gene pairs, but



interestingly, the A copy of this gene has been lost in both the teff and *E. pilosa* genomes, and a single loss of function allele would be sufficient for a mutant phenotype.

The hit on Chromosome 9B was also examined and there are many genes of interest within 100 Kb potentially involved in pigment synthesis: protein kinase domains, MYB-like DNA binding domains, HECT-domain, Acyl-CoA dehydrogenase, and Glycosyltransferase family 43 proteins. GO terms significantly associated with genes included involvement in steroid biosynthetic processes, potassium ion transmembrane transport, and metabolic processes.

## **Discussion and Conclusion**

Here, we developed a collection of teff representing substantial Ethiopian genetic diversity that will facilitate its utilization for commercial teff breeding. This unique study was aimed to dissect the USDA-GRIN germplasm and develop molecular breeding tools in teff. Sequencing of the TAP provides publicly available genotyping data that can be coupled with additional phenotyping to elucidate molecular markers and gene candidates for teff improvement with GWA and the employment of CRISPR-Cas9 gene editing.

We also provide an in depth study of the genetic diversity within the panel. Nine subpopulations were identified and they maintain Ethiopian genetic diversity by comparison to EtDP. Subpopulations 7 and 9 were more genetically diverse and clustered with the wild progenitor *E. pilosa* in contrast to the remaining seven subpopulations which had quite low FST scores. Nonetheless subpopulations exhibit a wide variability across phenotypes that can be explored. Considering the importance of seed color in the teff market, we chose to investigate 185 accessions for the genetic mechanisms underlying seed color variability. Seed color is a highly preferred trait by both growers and consumers. Although white seed is often favored, brown seed is sold at a lower price and thereby fills a hole in the market. Subpopulation 2 contains only white seeded individuals, and all other subpopulations have both white and brown seeded individuals; however, some have a stronger representation of a single color (Figure 1.4, Table 1.1). Subpopulations 1, 5, 6, 7, and 9 are majoritively brown seeded from 64.3-92.9%, while subpopulations 3, 4, and 8 are majoritively white seeded from 59.1-93.3%. The GWA of seed color provided a proof of concept, demonstrating the utilization of the TAP for genetic discovery. Additional phenotyping of the TAP for GWA to identify markers associated with agronomic and nutritional traits of importance will aid in the panel's utility.

Teff still displays a lack of traits we normally associate with domestication such as seed

shattering, seed size, and lodging tolerance. Although there are clear improvements from *E. pilosa*, there are still enhancements we can make relatively easily to rapidly advance teff varieties if we maintain and utilize the genetic diversity correctly. This suite of molecular resources that can be employed to improve future teff breeding strategies.

## **Methods**

### ***Plant Materials***

The majority of the germplasm analyzed in this study was sourced from the USDA-ARS Germplasm Resources Information Network (GRIN; <https://www.ars-grin.gov/>). This included accessions of *Eragrostis heteromera* (4), *Eragrostis macilenta* (1), *Eragrostis mexicana* (2), *Eragrostis papposa* (4), *Eragrostis pilosa* (13), and *Eragrostis tef* (363). Three accessions of *Eragrostis pilosa* were acquired from the Royal Botanical Gardens, KEW germplasm database.

For whole genome resequencing, each plant was cultivated in a separate 4-inch pot under controlled conditions with a 14-hour light and 10-hour dark cycle at temperatures of 26°C during the day and 20°C at night in a greenhouse setting. A single leaf segment (approximately 50 mg) was harvested from one plant per accession and immediately stored at -80°C for subsequent DNA extraction. Some traditional or farmer-maintained accessions of teff consist of mixed lines and from these, we selected a single representative plant for sequencing.

### ***DNA Extraction and DNA-Seq Library Construction***

For re-sequencing of teff accessions, DNA was extracted from leaves using the MagMax Plant DNA Isolation Kit (ThermoFisher # A32549). DNA concentration was measured using the Qubit HS DNA Kit (ThermoFisher # Q32854), and the quality of the DNA was verified by 0.8% agarose gel electrophoresis. Between 250-350 nanograms of DNA were used for DNA-seq library construction with the Kapa Hyper Plus DNA Kit (KapaBiosystems # KK8514), according to the manufacturer's protocol with 6 PCR cycles. Normalized, multiplexed DNA-seq libraries were pooled and sequenced on the HiSeq4000 system in paired-end 150 nt mode at the Michigan State University Genomics Core.

### ***Read alignment and variant detection***

Adapter sequences were trimmed from the paired-end reads using Trimmomatic v0.36. Reads shorter than 36 bp or with low-quality base pairs were removed. The trimmed reads were aligned to the *Eragrostis tef* genome assembly v3.1<sup>8</sup> using Bowtie2 v2.2.3 using default parameters. The mean read alignment rates for all teff accessions is 98.8%. Teff is an

allotetraploid, and the A and B subgenomes diverged an estimated ~5 million years ago. There is little evidence of homeologous exchange, and the A and B subgenomes have an average nucleotide similarity of 93%, and we observed proper read alignment to the A and B subgenomes. The resultant SAM files were sorted by chromosome, read group information was added, and converted to BAM format using Picard Tools v2.18.27. SNP calling was performed with GATK v3.8, adhering to the GATK Best Practice protocols. HaplotypeCaller was utilized to genotype each accession, and all resulting VCF files were merged into a single file using CombineGVCFs, followed by joint genotyping with GenotypeGVCFs. The final VCF file underwent filtering to remove InDels and SNPs with a depth of coverage (DP) less than 10 and a quality by depth (QD) less than 30 using GATK's SelectVariants function.

### ***Removing duplicated teff accessions***

To detect closely related or nearly identical accessions, we performed an identity-by-descent analysis using PLINK v1.9. SNPs within linkage disequilibrium blocks were first pruned using PLINK with the option `--indep-pairwise 50 10 0.5`. Paired accessions with a PI\_HAT value greater than 0.05 (indicating first to fourth degree relatives) were classified as cryptically related and grouped together. Within each cryptically related group, the accession with the highest number of sequencing reads was retained and the remaining accessions were removed from downstream analyses. All related accessions were identified exclusively within *Eragrostis tef* germplasm, with no related accessions found in other *Eragrostis* species or *Eragrostis pilosa*. In total, 135 accessions forming 36 groups were deemed related and 262 unique teff accessions were included for all downstream analysis.

### ***Nucleotide diversity estimation***

We estimated nucleotide diversity ( $\pi$ ) in teff and *E. pilosa* using the invariant sites of aligned reads to the teff reference genome using pixy (v1.2.7.beta)<sup>33</sup>. For this analysis, we reran variant calling using mpileup in bcftools (v1.9.64)<sup>34</sup> using the sorted bam files as described above, with default parameters and on each chromosome separately. The resulting VCF file contained read depth for each individual for every base pair of the genome, providing a framework to estimate nucleotide diversity more accurately. Using pixy, we calculated nucleotide diversity ( $\pi$ ) to estimate the genetic variation within each population, the fixation index (FST) to assess genetic differentiation among populations, and the average number of nucleotide substitutions per site between populations (Dxy) to understand the evolutionary

distances. We calculated these metrics for all teff accessions vs. *E. pilosa* or for each teff subpopulation separately.

### ***Locality analyses***

Longitude and latitude for the TAP were collected from passport data on the NPGS GRIN-GLOBAL website (<https://npgsweb.ars-grin.gov/gringlobal/search>). Where locality was not provided, longitude and latitude were estimated based on information listed. EtDP longitude and latitude coordinates were collected from previous publication<sup>9</sup>. Utilizing sf (<https://r-spatial.github.io/sf/>) a shapefile was created with longitude and latitude data and mapped to a shape file of Ethiopia from rnatualearth. Precipitation and elevation were obtained from geodata worldclim and raster then visualized in R version 4.3.3.

### ***Phenotyping***

The TAP was planted in triplicate at the Michigan State University Horticulture Teaching and Research Center Holt, MI (42°67'43.4"N, 84°48'43.5"W) in 2021 and 2022. Single row, 4.5 ft plots, were planted in a randomized complete block design. To enhance yield and uniformity, a fertilizer treatment of 19-19-19 at ~100 lbs/A was applied prior to planting. Additionally, herbicide was applied to the entire field for the control of broadleaves (Broclean), and between rows for grasses (Roundup PowerMAX). Seed was harvested, threshed, and cleaned, then classified as brown, white, or mixed seed.

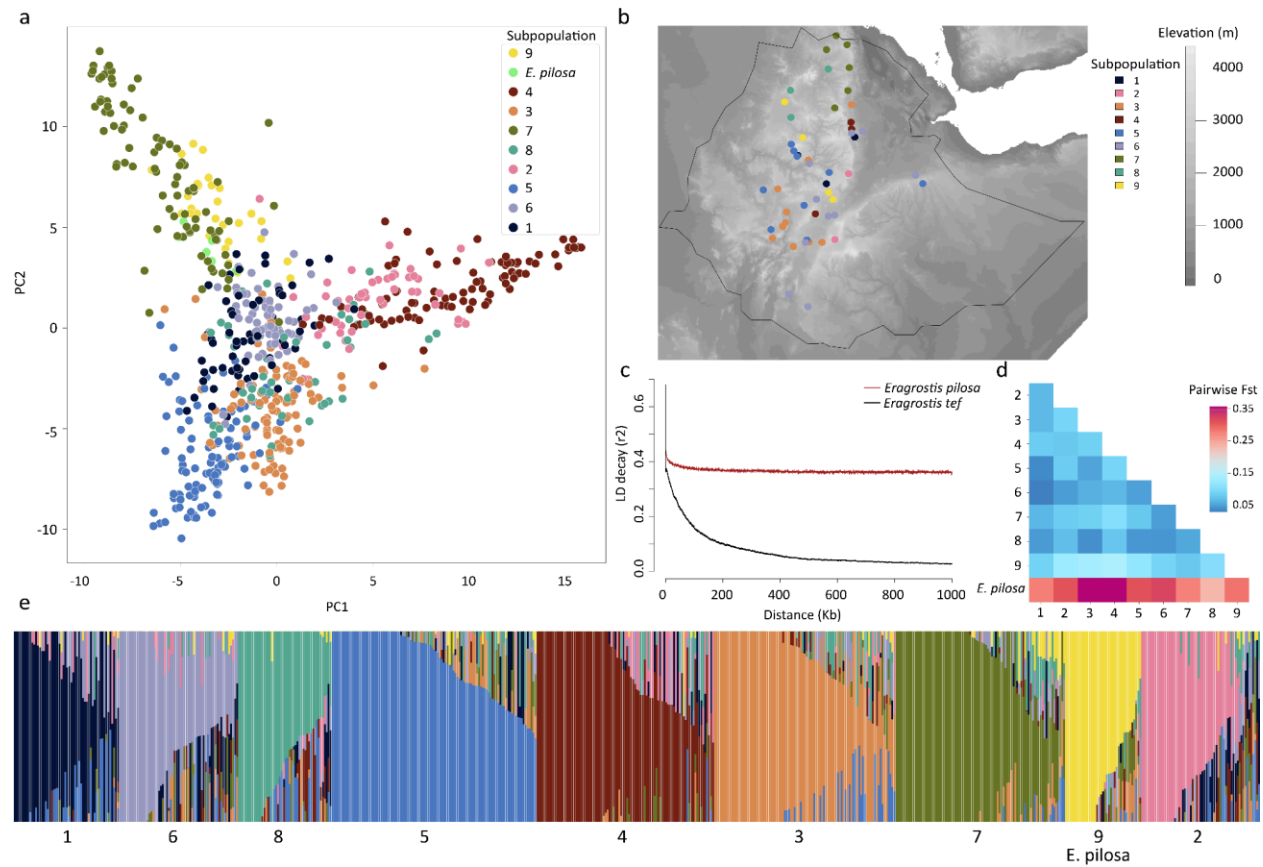
### ***Putative Seep Identification:***

Using XPCLR, we identified selective sweeps from 8.1 million SNPs. XPCLR was calculated with a 50-Kb sliding window and 25-Kb step size using the updated software<sup>35</sup>. The top one percent of XPCLR values were considered candidate regions.

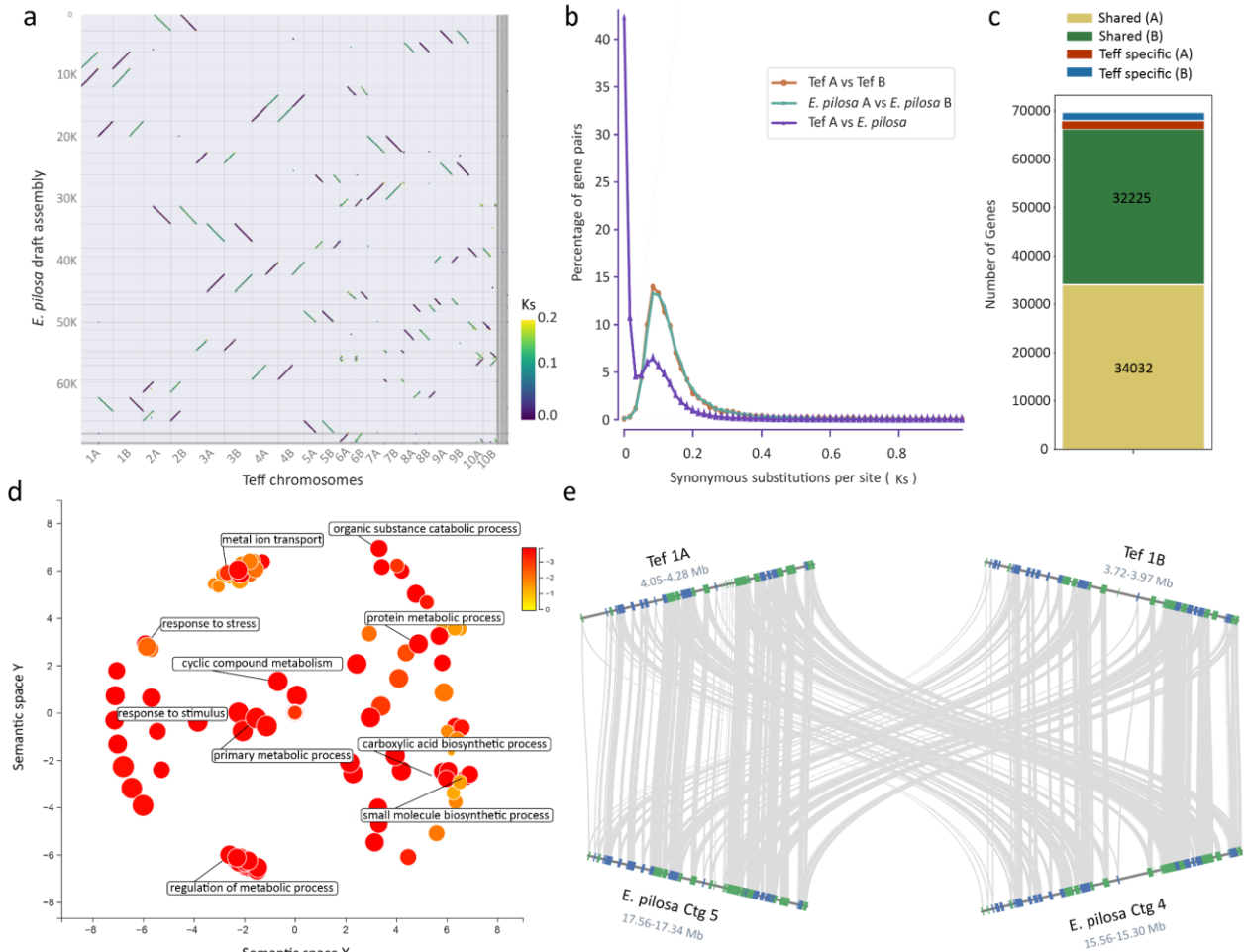
### ***Genome wide association***

Samples with consistent seed color of white and brown were selected for further analysis. Genome wide association was performed on 185 accessions using Bayesian-information and LD iteratively nested keyway (BLINK) in GAPIT version 3 in R<sup>36,37</sup>. A kinship matrix and the first three PCs of the TAP PCA were included as covariates. Manhattan plots and QQ plots were constructed using QQman<sup>38</sup>. Single nucleotide polymorphisms with a pvalue greater than Bonferroni corrected value were selected as significantly associated loci. Genes 100 Kb upstream and downstream of the loci were evaluated as putative candidates and were reviewed with associated GO terms (Table 1.2).

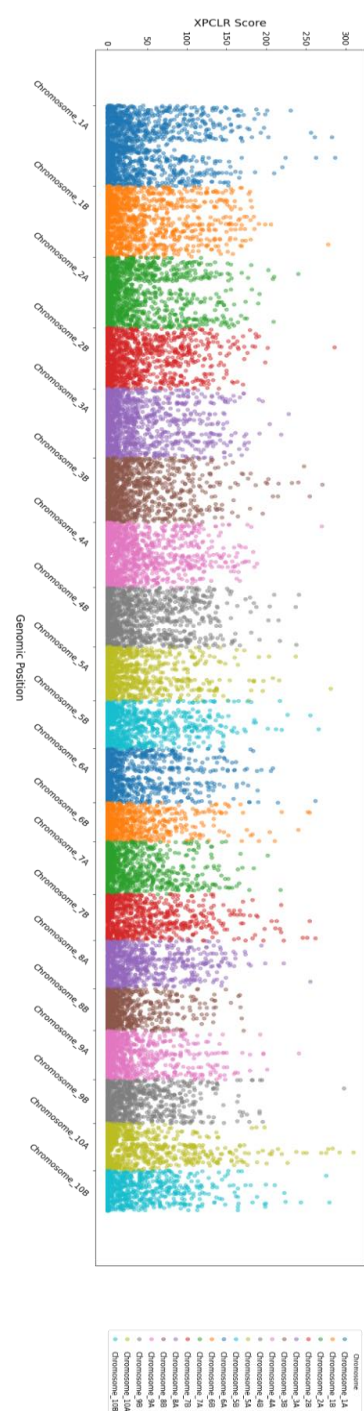
## Tables and Figures



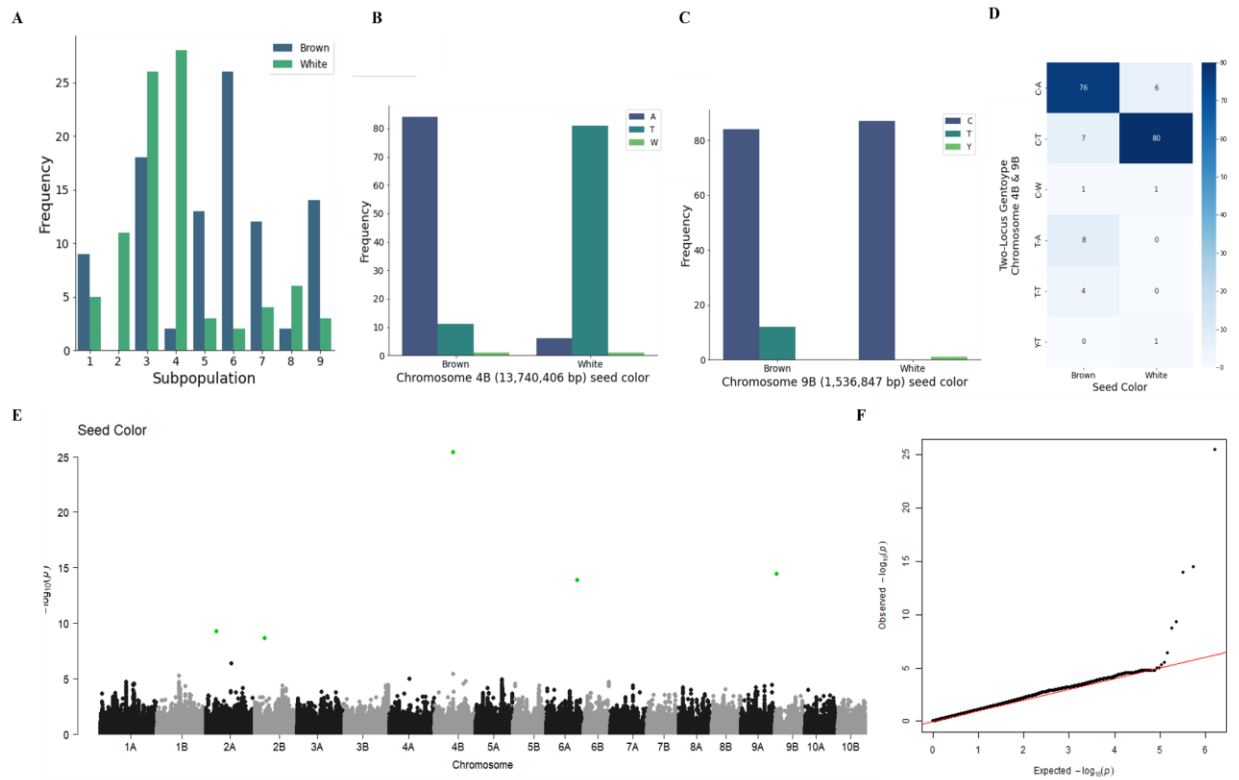
**Figure 1.1: Genetic diversity of teff and its wild progenitor *Eragrostis pilosa*.** (a) Principal component analysis of commonly genotyped SNPs between the EtDP and TAP lines. Accessions are colored by the nine subpopulations identified in the ADMIXTURE analysis. The *E. pilosa* accessions are found in subpopulation 9, but are colored separately here. (b) Distribution of georeferenced lines from the TAP across an altitudinal map of Ethiopia. (c) Linkage disequilibrium decay plot for *E. pilosa* and only teff lines from the TAP using the full set of genome wide variants. (d) Pairwise population fixation (Fst) between accessions in each of the 9 teff subpopulations and *E. pilosa*. (e) ADMIXTURE results for the common SNPs between the EtDP and TAP. Each accession is represented by a vertical line, and colored by the proportion of each subpopulation in the genome of each line.



**Figure 1.2: Comparative genomics of the *teff* and *E. pilosa* genomes.** (a) Macrosyntentic dot plot between the *E. pilosa* and *teff* genomes where each dot represents a syntenic gene pair and dots are colored by the  $K_s$ . (b) Histogram of  $K_s$  for syntenic gene pairs between *E. pilosa* and *teff* (purple), homeologs between the *teff* A and B subgenomes (orange) or *E. pilosa* A and B subgenomes (blue). (c) Stacked bar plot showing the gene pairs conserved between *teff* and *E. pilosa* and genes unique to *teff* for the A and B subgenomes. (d) Enriched Gene Ontology (GO) terms of genes that are specific to *teff*. GO terms are transformed using Multidimensional Scaling to reduce dimensionality and terms are grouped by semantic similarities. The color and size of the circles represent significance, and clustered processes of interest are highlighted. (e) Microsynteny between the *teff* and *E. pilosa* genomes. A portion of Chromosomes 1A and B are shown where individual genes are shown in blue or green and syntenic gene pairs are connected by gray lines.



**Figure 1.3: XPCLR hits across the teff genome.**



**Figure 1.4: Seed color across the TAP.** A) Seed color distribution across subpopulations. B) Allelic distribution of phenotypic variation at Chromosome 4B (13,740,406 bp). C) Allelic distribution of phenotypic variation at Chromosome 9B (1,536,847 bp). D) Heatmap of genotypic combinations for Chromosome 4B (13,740,406 bp) and Chromosome 9B (1,536,847 bp). E) BLINK GWA Manhattan plot, significant loci highlighted in green. F) QQ plot for GWA.



**Table 1.1:** Percentage of brown and white seeded teff in each subpopulation included in GWA with 185 accessions.

Subpopulation	Percentage of Brown Seed	Percentage of White Seed
1	64.3%	35.7%
2	0.00%	100%
3	40.9%	59.1%
4	6.67%	93.3%
5	81.3%	18.8%
6	92.9%	7.10%
7	75.0%	25.0%
8	25.0%	75.0%
9	82.4%	17.6%

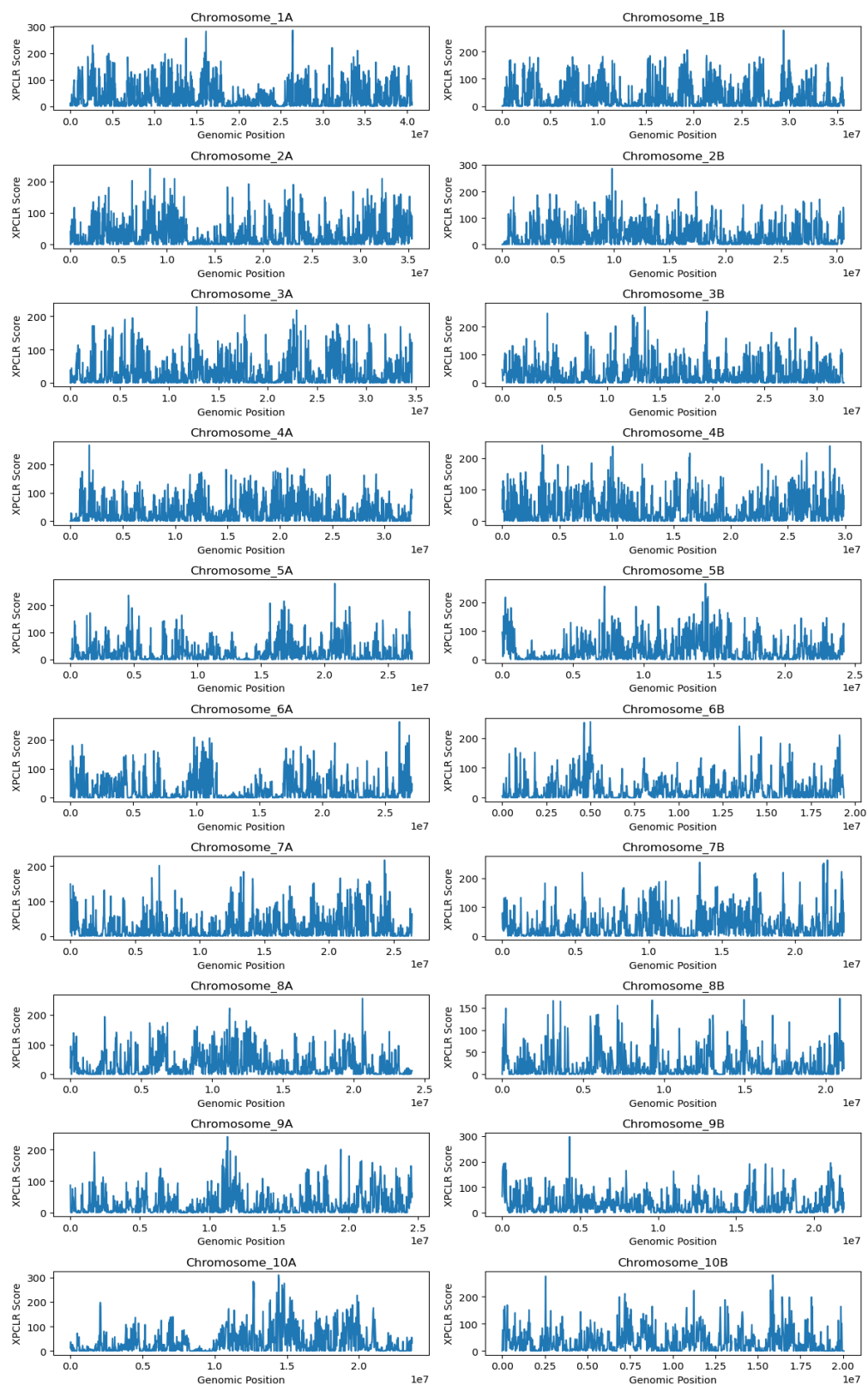
**Table 1.2:** Significantly associated GO terms of genes +/- 100kb of loci from GWA.

GO.ID	Term	Classic Fisher	Gene	SNP
GO:0006694	steroid biosynthetic process	0.022	Et_4B_037027	Chromosome_4B_13795495
GO:0071805	potassium ion transmembrane transport	0.043	Et_2A_017894	Chromosome_2A_7583449
GO:0008152	metabolic process	0.059	Et_2A_017889	Chromosome_2A_7583449
GO:0008152	metabolic process	0.059	Et_2B_021964	Chromosome_2B_7457292
GO:0008152	metabolic process	0.059	Et_2B_021976	Chromosome_2B_7457292
GO:0008152	metabolic process	0.059	Et_2B_022890	Chromosome_2B_7457292

**Table 1.2 (cont'd)**

GO:0008152	metabolic process	0.059	Et_6A_046928	Chromosome_6A_22495014
GO:0008152	metabolic process	0.059	Et_9B_064420	Chromosome_9B_1536847
GO:0008152	metabolic process	0.059	Et_9B_064507	Chromosome_9B_1536847

## Supplement:



**Figure Supplemental 1: XPCLR results for each chromosome.**

## REFERENCES

1. Minten, B., Taffesse, A. S. & Brown, P. *The Economics of Teff: Exploring Ethiopia's Biggest Cash Crop*. (Intl Food Policy Res Inst, 2018).
2. Csa. Federal Democratic Republic of Ethiopia: Central Statistical Agency: Agricultural Sample Survey. Preprint at (2021).
3. Reda, Abraham. Achieving Food Security in Ethiopia by Promoting Productivity of Future World Food Tef: A Review. *Advances in Plants & Agriculture Research* **2**, (2015)
4. Bachewe, F. N., Koru, B. & Taffesse, A. S. Productivity and efficiency of smallholder teff farmers in Ethiopia. *Gates Open Res* **3**, 208 (2019).
5. Tadele, E. & Hibistu, T. Empirical review on the use dynamics and economics of teff in Ethiopia. *Agriculture & Food Security* **10**, 1–13 (2021).
6. Assefa, K., Chanyalew, S. & Tadele, Z. Tef, *Eragrostis tef* (Zucc.) Trotter. in *Millet and Sorghum* 226–266 (John Wiley & Sons, Ltd, Chichester, UK, 2017).
7. Beyene, G. *et al.* CRISPR/Cas9-mediated tetra-allelic mutation of the 'Green Revolution' SEMIDWARF-1 (SD-1) gene confers lodging resistance in teff (*Eragrostis tef*). *Plant Biotechnol. J.* **20**, (2022).
8. VanBuren, R. *et al.* Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
9. Woldeyohannes, A. B. *et al.* Data-driven, participatory characterization of farmer varieties discloses teff breeding potential under current and future climates. *Elife* **11**, (2022).
10. Baye & Kaleab. *Teff: Nutrient Composition and Health Benefits*. (Intl Food Policy Res Inst, 2014).
11. Cheng, A., Mayes, S., Dalle, G., Demissew, S. & Massawe, F. Diversifying crops for food and nutrition security--a case of teff. *Biol. Rev. Camb. Philos. Soc.* **92**, 188–198 (2017).
12. Engels, J., Hawkes, J. G. & Worede, M. *Plant Genetic Resources of Ethiopia*. (Cambridge University Press, 1991).
13. Demissie, A. Tef genetic resources in Ethiopia. *on Tef Genetics and Improvement, Debre Zeit, Ethiopia*.
14. D'Andrea, A. C. T'ef (*Eragrostis tef*) in Ancient Agricultural Systems of Highland Ethiopia. *Econ. Bot.* **62**, 547–566 (2008).
15. Costanza, S. H., Dewet, J. M. J. & Harlan, J. Literature review and numerical taxonomy of *Eragrostis tef* (T'ef). *Econ. Bot.* **33**, 413–424 (1979).

16. Tefera, H., Assefa, K. & Belay, G. Evaluation of interspecific recombinant inbred lines of *Eragrostis tef* x *E. pilosa* [Ethiopia]. *J. Genet. Breed.* (2003).
17. Girma, D., Cannarozzi, G., Weichert, A. & Tadele, Z. Genotyping by Sequencing Reasserts the Close Relationship between *Tef* and Its Putative Wild *Eragrostis* Progenitors. *Diversity* vol. 10 17 Preprint at <https://doi.org/10.3390/d10020017> (2018).
18. Ingram, A. L. & Doyle, J. J. The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear waxy and plastid rps16. *Am. J. Bot.* **90**, 116–122 (2003).
19. Jones, B. M. G., Ponti, J., Tavassoli, A. & Dixon, P. A. Relationships of the Ethiopian Cereal T'ef (*Eragrostis tef* (Zucc.) Trotter): Evidence from Morphology and Chromosome Number. *Ann. Bot.* **42**, 1369–1373 (1978).
20. Tao, Y. *et al.* Extensive variation within the pan-genome of cultivated and wild sorghum. *Nat Plants* **7**, 766–773 (2021).
21. Zhao, Q. *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **50**, 278–284 (2018).
22. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
23. VanBuren, R., Bryant, D., Edger, P. P., Tang, H. & Burgess, D. Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeum*. *Nature* (2015).
24. VanBuren, R., Wai, C. M., Keilwagen, J. & Pardo, J. A chromosome-scale assembly of the model desiccation tolerant grass *Oropetium thomaeum*. *Plant Direct* (2018).
25. Marks, R. A., Van Der Pas, L., Schuster, J. & VanBuren, R. Convergent evolution of desiccation tolerance in grasses. *bioRxiv* 2023.11.29.569285 (2023) doi:10.1101/2023.11.29.569285.
26. Pardo, J. *et al.* Intertwined signatures of desiccation and drought tolerance in grasses. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10079–10088 (2020).
27. Adnew, T., Ketema, S., Tefera, H. & Sridhara, H. Genetic Diversity in *Tef* [*Eragrostis tef* (Zucc.) Trotter] Germplasm. *Genet. Resour. Crop Evol.* **52**, 891–902 (2005).
28. Zhou, Y. *et al.* Triticum population sequencing provides insights into wheat adaptation. *Nat. Genet.* **52**, 1412–1422 (2020).
29. Wu, D. *et al.* Genomic insights into the evolution of *Echinochloa* species as weed and orphan crop. *Nat. Commun.* **13**, 689 (2022).
30. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).

31. Peer, W. A. *et al.* Flavonoid Accumulation Patterns of Transparent Testa Mutants of *Arabidopsis*. *Plant Physiol.* **126**, 536 (2001).
32. Schoenbohm, C., Martens, S., Eder, C., Forkmann, G. & Weisshaar, B. Identification of the *Arabidopsis thaliana* Flavonoid 3'-Hydroxylase Gene and Functional Expression of the Encoded P450 Enzyme. **381**, 749–753 (2000).
33. Korunes, K. L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
34. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
35. GitHub - xuzhougeng/xpclr: Code to compute the XP-CLR statistic to infer natural selection. *GitHub* <https://github.com/xuzhougeng/xpclr>.
36. Wang, J. & Zhang, Z. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics Proteomics Bioinformatics* **19**, 629–640 (2021).
37. Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **8**, giy154 (2018).
38. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* (2014) doi:10.1101/005165.

### **CHAPTER 3: GENOME WIDE ASSOCIATION OF LODGING SUSCEPTIBILITY AND PANICLE ARCHITECTURE TRAITS IN TEFF**

The work presented in this chapter is being prepared for publication:

Wilson ML, Smith-Corso I, Zahra C, Thompson A, VanBuren R

Genome wide association of lodging susceptibility and panicle architecture traits in teff. *In prep.*

Author contributions: MLW and RV designed the experiments. MLW completed all experiments and data analysis. AT provided expertise and guidance on data analysis and interpretation. MLW wrote the manuscript. RV edited the manuscript.

## Abstract

Teff (*Eragrostis tef*) is a climate resilient grain crop most commonly grown by small-scale Ethiopian farmers. Although economically and culturally important to the Horn of Africa, the crop is globally underutilized due to its lower yield when compared to other cereals. However, teff has immense genetic diversity that can be harnessed for rapid improvement with strategic breeding. Lodging is a major contributor to low yields in teff but poses a big challenge to breeders because it is a quantitative trait impacted by morphology, environment, and stress response. We surveyed the Teff Association Panel (TAP) consisting of 259 *Eragrostis tef* and 6 *Eragrostis pilosa* USDA-GRIN accessions at Michigan State University in 2021 and 2022 for lodging susceptibility, panicle architecture, plant height, days to heading, culm width, average panicle weight, and average seed weight per panicle. The association between traits were evaluated to determine which traits play a critical role in lodging tolerance and susceptibility in teff. Panicle architecture is highly correlated with lodging susceptibility, and accessions with open panicle architectures lodged more consistently. The phenotypic data was leveraged to perform genome wide association (GWA) for each trait and identify 26 loci. These loci and putative gene candidates may facilitate the future development of teff cultivars to strengthen lodging tolerance and preserve plant height and yield.

## Introduction

Teff was domesticated in the highlands of Ethiopia, where selection favored stability and consistent harvest over yield. Although understudied in the Global North and often referred to as an underutilized or orphan crop, teff serves as an essential staple grain, providing food security and economic opportunities for farmers and growers across the Horn of Africa. Ethiopia is the top teff producer, where the crop is traditionally grown by smallholder farmers and used to prepare injera, a fermented flatbread that is consumed daily. In 2006, the Ethiopian government placed an export ban on teff to address the rising market cost and sustain future food security, thereby stimulating the industry around the world <sup>1</sup>. In the United States, teff production as an alternative grain and forage crop has grown exponentially <sup>2,3</sup>. Although generally lower yielding than leading cereals, teff is adapted to dry conditions unsuitable for other grain crops, making it an attractive alternative grain for production in low rainfall areas.

Despite its climate resilience, teff generally lacks agronomic improvement of traits we associate with domestication, such as lodging tolerance, increased seed size, and plot uniformity



<sup>4</sup>. There is a wealth of traditional knowledge in teff, but breeding efforts have lagged behind other cereals because of the difficulty in crossing teff and limited scientific funding for teff breeding and production research. Despite these limitations, tremendous phenotypic and genetic diversity has been maintained indigenously across thousands of locally adapted farmer varieties and breeding lines of teff at the Ethiopian Institute of Agricultural Research. Participatory varietal selection with the involvement of local Ethiopian farmers was recently conducted to evaluate the comprehensive Ethiopian Teff Diversity Panel (EtDP). Farmers prefer tall, high yielding, high biomass, and fast maturing lines with long, open panicles regardless of their genetic background <sup>5</sup>. With a focus on Ethiopian farmer and consumer preference as well as more accurate screening, high-throughput phenotyping, and genetic resources, rapid trait improvement while maintaining the rich genetic diversity and local adaptation of teff can be achieved.

Lodging is the largest limitation of teff yield <sup>6</sup>. Teff experiences both root and stem lodging, where root lodging disrupts the plant's anchorage, while stem lodging compromises the stalk's structural integrity. <sup>7, 8</sup>. This condition is triggered by stress at the stem or root, causing the tiller to collapse at an angle, which may lead to a domino effect where adjacent tillers lean on each other, pushing more panicles toward the soil. In Ethiopia, lodging results in a massive decrease in yield ranging from 20%–29% of teff fields on average <sup>9–11</sup>. Lodging makes harvest more difficult, both manually and mechanically, and as panicles reach the soil they become more susceptible to disease and pests as well as germination following enough rain. Because it is an environmentally affected trait with increased susceptibility from wind and rain, lodging is incredibly difficult to phenotype and genetically characterize. Lodging is also influenced by various aspects of plant morphology including plant height, seed filling, plant density, tillering, root-to-shoot ratio, stem strength and width, panicle length, and panicle weight, making it a complex trait to manage and study.

To develop a breeding strategy for lodging tolerance in teff, understanding variance of lodging susceptibility and morphological traits associated with tolerance is essential. A genome wide association study previously identified six loci associated with lodging index across water limited and well-watered environments <sup>9</sup>. The researchers expanded on lodging loci with strong pleiotropic effects involving grain yield and maturity but did not identify putative candidate genes associated with these marker–trait associations.

Lodging susceptibility can be decreased by dwarfing or optimizing agronomic practices such as nitrogen fertilization and sowing rate, but many varieties remain at the mercy of the weather due to their morphology <sup>11</sup>. Improving cereals through breeding for shorter varieties and identifying dwarfing genes has successfully reduced lodging in rice, wheat, and maize. Recent advancements in teff breeding and genetics have facilitated the establishment of semi-dwarf RILS and gene-edited cultivars with decreased height <sup>12,13</sup>. Previous studies have shown that panicle angle and peduncle-panicle length are correlated with teff lodging tolerance <sup>14,15</sup>. Although this is an outstanding improvement in teff lodging research, we hypothesize that lodging tolerance can be further improved by focusing on panicle morphological traits.

Here, we searched for associations of lodging tolerance and various plant architectural and agronomic traits in teff using the recently established Teff Association Panel (TAP). The TAP encompasses a representative collection of germplasm from major teff producing regions in Ethiopia as well as the wild progenitor of teff, *Eragrostis pilosa*. We report the prevalence of lodging across the panel and highlight significant associations with traits including panicle architecture, height, and panicle weight. Finally, we identified specific genetic loci through genome-wide association (GWA) that offer potential targets for further validation and integration in teff breeding programs.

## **Methods**

### ***Plant materials and field conditions***

We surveyed lodging and plant agronomic traits using the Teff Association Panel (TAP), which includes representation from the nine subpopulations of teff across the major growing regions in Ethiopia as well as the wild progenitor of teff, *Eragrostis pilosa*. Accessions within the TAP are available from the USDA germplasm Resources Information Network (GRIN) and have been fully resequenced. The panel was grown in triplicate using a randomized block design in the summer of 2021 and 2022 with single row, 4.5 ft plots, at the Michigan State University Horticulture Teaching and Research Center (HTRC) in Holt, MI (42°67'43.4"N, 84°48'43.5"W). Approximately 100 teff seeds of each accession were planted by hand in each plot on June 1st in 2021 and June 3<sup>rd</sup>, 2022. To enhance yield and uniformity, a fertilizer treatment of 19-19-19 at ~100 lbs/A was applied prior to planting. Additionally, herbicide was applied to the entire field for the control of broadleaves (Broclean), and between rows for grasses (Roundup PowerMAX). Soil health is maintained by the HTRC, and soil cores were sampled at the end of each season

across the field and presented optimum phosphorus, potassium, and magnesium levels with slightly high pH of 6.8 and 7.6 on average in 2021 and 2022. Plots were harvested by hand November 2nd-10th, 2021 and October 20-28th, 2022.

### ***Phenotyping of lodging, architecture, and agronomic traits***

In 2021 and 2022 data for four and six traits respectively, were collected related to lodging, plant architecture, and yield. Plant height was measured post flowering, but before grain filling, using an average from two mid-row plants measured from ground to the maximum height of the panicle. At maturity, each plot was scored for panicle architecture using the scale of 1-4 previously developed, where 1= very compact, 2=semi-compact, 3=fairly loose, and 4= very loose <sup>16,17</sup>. Heading date was collected when more than half of the plants in each plot had visible panicles and subtracted from the planting date of June 1st and June 3rd in 2021 and 2022, respectively. Differences in weather patterns and rainfall led to significant variation in the distribution of heading date between years, and we divided accessions into early, mid and late heading to enable better comparisons. Panicle architecture traits were measured using three representative panicles that were selected from mid-row plants in each plot and collected near the end of the grain filling period. Fully dried panicles were imaged separately and then weighed together to get an average panicle weight. The three panicles were then threshed and the seed from each set of three panicles was pooled and weighed together then averaged for each plot. Lodging severity was measured before harvest using a severity score ranging from 0 to 5. This scale quantifies the percentage of panicles that had lodged within the plot where a score of 0 indicates 0% lodging, with all panicles erect; 1 signifies that up to 20% of panicles had lodged; 2 represents 20-40% lodging; 3 for 40-60%; 4 denotes 60-80%; and 5 corresponds to 80-100% of the panicles lodged.

### ***Statistical analysis of field data***

Statistical analyses were performed in R software version 4.3.3 unless otherwise noted. To account for the skew of categorical variables, lodging and panicle architecture scores were transformed using boxcox in the MASS package <sup>18</sup>. An ANOVA was carried out to model the genetic, block, year, and subpopulation effects on each trait. The Best Linear Unbiased Estimators (BLUES) for each trait were estimated using lme4 where block and year were fit as random effects and accession was fit as a fixed effect for each trait, since plots were planted in a randomized complete block design. BLUES of each accession for each trait were used as

phenotype data for GWA. Broad-sense heritability ( $h^2$ ), or in this case repeatability, was estimated as a secondary model in which accession, block, and year were fit as random effects to estimate variance components via restricted maximum likelihood (REML) using lme4.  $H^2$  values of the traits were estimated according to  $h^2 = \frac{\sigma^2}{\sigma^2 + \sigma e^2}$ , where  $\sigma^2$  is the genotypic variance and  $\sigma e^2$  is the residual variance. Probit ordinal regression was conducted for each year using scipy.stats (statsmodels v0.14.2) <sup>19</sup> with lodging score as a categorical variable and panicle architecture score, height, days to heading, panicle weight, and seed weight as predictors. Pearson's correlation of BLUES of each trait was calculated using the corr() function in the Python package pandas (v2.2) <sup>20</sup>. Lastly, a trait PCA biplot was produced via sklearn evaluating trends across trait BLUES and subpopulation.

### ***Genome wide association***

Whole genome resequencing data was previously generated for the TAP, and we used a subset of LD pruned single nucleotide polymorphisms as genetic markers for genome wide association (GWA). In total, 746,242 LD pruned, imputed SNP based markers were used. A kinship relatedness matrix and principal component analysis on genetic structure of the panel were conducted in TASSEL <sup>21</sup>. The three principal components and kinship matrix were applied as covariates to account for population structure and relatedness of the panel. GWA for lodging, panicle architecture, height, days to heading, panicle weight, and seed weight were conducted using GAPIT version 3.4 in R with trait BLUES <sup>22</sup>. GWA was performed using a range of models including general linear model (GLM), mixed linear model (MLM), compressed MLM (CMLM), multiple loci mixed model (MLMM), Fixed and random model Circulating Probability Unification (FarmCPU), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK). Model performance was evaluated using quantile-quantile plots and based on model fit, the FarmCPU model was chosen for further analysis. P-values for each SNP were visualized as Manhattan plots with QQMan <sup>23</sup>. Candidate genes for each trait were identified based on their proximity of 100 kb up- and downstream of significant markers identified in the GWA. Candidate genes were annotated based on KEGG and GO terms to describe generalized metabolic and cellular processes responses. KEGG annotations were generated for each gene using BLASTKoala (<https://www.kegg.jp/blastkoala/>) and were used to create metabolic pathway maps with KEGGmapper (<https://www.genome.jp/kegg/mapper/color.html>).

## Results

### *Teff Association Panel phenotypic diversity*

To investigate the association of lodging and plant architecture traits in teff, we grew the TAP in 2021 and 2022 in East Lansing Michigan and collected a range of phenotypic data across the growing season. The teff association panel contains a broad diversity of agronomic traits including plant height, days to heading, panicle architecture, culm width, panicle weight, and average seed weight per panicle (**Fig. 2.1**). Days to heading was the least variable trait across the panel with a coefficient of variation (CV) of 16.58, followed by height, culm width and panicle architecture with CV of 25.63, 31.87, and 36.02, respectively (**Table 2.1**). Lodging, panicle weight, and seed weight were highly variable across the TAP (CV= 68.93, 61.03, and 94.97 respectively). Cumulative rainfall was significantly higher in 2021 compared to 2022, and this precipitation promoted faster maturation as well as increases in plant biomass, height, and lodging in 2021 (**Supplemental Fig. 2.1**). To test the significance of genotypic and environmental effects, we performed ANOVA for each trait. Year was significant across each trait where we collected two years of data including lodging, panicle architecture, height, and days to heading (**Table 2.2**). Block was highly significant for lodging as well as the single year traits of panicle weight, seed weight, and culm width (**Table 2.2**). Highly significant block effects for lodging severity may have resulted from wind and rain direction. Block significance may also stem from variability in soil characteristics across the field. Genotypic effects were significant for all traits except for culm width (**Table 2.2**). The lack of a significant genotype effect in culm width is likely due to the narrow distribution of width values across the panel. Each trait we measured across the TAP had high broad-sense heritability, or repeatability in the field, ranging from 0.63 for height to 0.93 for panicle architecture, except for culm width which had a low heritability of 0.15 (**Table 2.1**). These high heritabilities for the majority of traits highlight the genetic variance within the panel that can be harnessed in future breeding efforts.

### *Association between lodging tolerance and plant architecture*

Lodging was prevalent in 2021 and 2022 across the surveyed teff germplasm. Overall lodging severity was higher in the 2021 field season, with 91% of plots having a lodging score of one or higher, corresponding to > 20% of plants lodging, compared to 80% of the plots in 2022 having lodging scores greater than or equal to one (**Fig. 2.1**). Lodging scores were higher for 87% of teff accessions in 2021 compared to 2022. Despite the prevalence of lodging, we

identified 8 teff and 1 *E. pilosa* accession that did not lodge within two or more replicated blocks in both years: PI405074, PI494384, PI494234, PI329681, PI494459, PI494313, PI494331, PI194924, and PI219588 (*E. pilosa*). These accessions have mostly compact or semi-compact panicles and broad trait variability in height ranging from 7.78-131.92 cm, panicle weights ranging from 0.04-2.70 grams, and corresponding seed weights ranging from 0.003-1.26 grams.

Lodging tolerance is significantly influenced by several plant architecture and developmental traits in teff. Lodging severity is negatively correlated with plant height ( $r=-0.68$ ,  $pval=8.08e-38$ ), days to heading ( $r=-0.43$ ,  $pval=3.13e-06$ ), panicle weight ( $r=-0.46$ ,  $pval=3.98e-36$ ) and seed weight ( $r=-0.18$ ,  $pval=4.43e-16$ ) and positively correlated with panicle architecture score ( $r=0.57$ ,  $pval=1.68e-39$ ) (**Fig. 2.2**). These correlations suggest that accessions with low lodging are generally high yielding and tall, with more compact panicle architecture and a later flowering time. To investigate this further, we ran probit ordinal regression with lodging as a categorical variable and traits as predictors (**Supplemental Table 2.1**). Due to variation, each year was run separately. For data collected in 2021, the regression analysis predicts a 0.3698 unit increase in lodging severity for every increase of panicle architecture score, translating to a 37% increase in lodging score for each category of increasing panicle openness. Conversely, for each additional centimeter in plant height, millimeter in culm width, and day to flowering, lodging severity is reduced by 0.03, 0.04, and 0.04 units respectively. In 2022, more open panicles maintained the increase in lodging severity by 37%, but the trend for height is reversed, and lodging severity increases by 0.02 units every additional centimeter of height. We also see a very strong predictor in panicle weight in which lodging score decreases by 1 for every gram increase in panicle weight. It seems contradictory that height is negatively correlated with lodging in teff, but shorter plants had more open panicle architecture and lodged more consistently within TAP, and taller plants had a higher proportion of accessions with compact architecture and they did not lodge as consistently. This trend between plant height and lodging susceptibility shifts for the 2022 field season, but the mean height of plants also decreased by 5 cm. We compared our height data to the USDA GRIN website and found that the plants grown in our experiment, across both years, were not as tall as field trials in Pullman WA (**Supplemental Fig. 2.2**). Our fields were not irrigated and experienced highly variable precipitation, had little fertilizer, and we used different field management strategies. Nonetheless, shorter plants with lighter panicles are not always more lodging tolerant, and farmer-preferred loose panicle architectures are

generally more susceptible to lodging.

### ***Association of traits across domestication and teff subpopulations***

We previously completed whole-genome resequencing of TAP and identified nine subpopulations of teff. To understand and dissect the genetic architecture of desirable traits within the panel, we ran a principal component analysis (PCA). The first three components explained approximately 17% of the global genetic variation in TAP. Even with the 13% variance explained in the first two PCs we see clear segregation of subpopulations (**Fig. 2.3**). To explore phenotypic diversity across the subpopulations, we calculated best linear unbiased estimators (BLUES) to adjust for environmental variability and minimize bias, and compared BLUE scores across subpopulations for each trait (**Fig. 2.4**). We also evaluated traits in wild *Eragrostis pilosa* accessions to establish phenotypes of traits prior to domestication. We see variability in distributions for each subpopulation across traits. Coefficients of variation for days to heading was relatively low across all subpopulations from 13-19% (**Table 2.3**). Height also had low coefficients ranging from 20-27 except for *E. pilosa*, which had a coefficient of 50. *E. pilosa* has low variation in panicle architecture compared to the rest of the subpopulations, as all surveyed wild accessions have very open panicles. *E. pilosa* had a mean lodging score of 2, very open panicles, shorter plant height, with earlier flowering times, smaller culms, and lower panicle and seed weights. Subpopulations 7 and 9 have the lowest genetic distance and population differentiation ( $F_{st}$ ) values when compared to *E. pilosa*, and we see similar trends across phenotypes. Subpopulations 7 and 9 have shorter plants, small culm widths, and lower seed and panicle weights compared to other teff subpopulations, however, they have more variability in lodging and panicle architecture score as well as maturity.

We fit a linear model to estimate significance of the effect of subpopulation across traits. Statistically significant differences in height, panicle weight, seed weight, lodging, and panicle architecture were observed across subpopulations. Subpopulations 4, 7, and 8 have significant differences in height, and subpopulations 2, 4, and 8 exhibited significant differences in panicle weight. Subpopulation 4 was the only significant subpopulation for seed weight, and subpopulations 2, 4, 7, and 8 were highly significant in panicle architecture. Subpopulations 3 and 5 have significantly different lodging susceptibility. Subpopulation 4 exhibited significantly different phenotypes across four traits: height, panicle weight, seed weight, and panicle architecture. This subpopulation had generally tall plants with a mean height of 66cm, large

panicles with the highest seed weight, and compact panicle architectures with a mean score of ~2. Subpopulation 8 follows a similar trend with the tallest plants at a mean of 70 cm, heavy panicles, and an average panicle architecture score of 1.83. However, there are only 11 accessions in this subpopulation, which contributes to the lack of variability and heightened statistical significance. Passport data is minimal on the USDA GRIN website, and the majority of lines are recorded from Debre Zeit Experimental Station as well as markets across Asmara, Addis Ababa, or Dire Dawa in subpopulation 4 and Gondar in subpopulation 8, so we cannot speculate about local adaptation. Subpopulation 4 contains an improved variety, ‘Magna’, as well as a landrace, ‘Gommadie’, and market cultivars Manjna and Sergeyna. The cultivar Addisie is in subpopulation 8.

Lodging susceptibility is significantly different in subpopulation 3 and 5. Subpopulation 3 is the most populated with 56 accessions and has the highest proportion of fully lodged and fully erect plots. Subpopulation 5 ranks second in severe lodging score majority. Lodging resistant phenotypes were found in subpopulations 2, 3, 4, 5 and 8. When we compare to panicle architecture, subpopulations 3 and 8 have the most compact panicles while subpopulation 2 has no accessions with fully compact panicles, or a score of 1.

To analyze patterns of associated traits, we conducted a principal component analysis on the BLUEs and visualized the results in a biplot. Principal components 1 and 2 explain 46% and 24% of the variance respectively (**Fig. 2.5**). We also evaluated the relative contribution of traits to the variation observed among subpopulations. Populations 1-5 and populations 6-9 plus *E. pilosa* cluster together along PC1 and PC2. Notably, populations 1-5 exhibited less consistent lodging, more compact panicle architecture, taller stature, and greater seed yield.

### ***Identifying genetic loci associated with agronomic traits***

We conducted genome-wide association (GWA) using multiple models implemented in GAPIT to identify genetic loci underlying agronomic traits across the 259 teff accessions in the TAP. The FarmCPU models consistently showed better fit in QQ plots for each trait assessed, so this model was used for all GWA. FarmCPU employs both mixed linear model and general linear model strategies to compare results, and this approach helps determine whether markers significantly influence the phenotype while accounting for population structure. Therefore, FarmCPU produces less background noise and identifies a selection of fewer loci that have strong, significant effects.



A total of 26 significant loci were detected across the six traits of plant height, days to heading, panicle architecture, lodging, panicle weight, and average seed weight per panicle (**Table 2.4**). Seven loci were detected for both height and lodging susceptibility, followed by four for panicle architecture and seed weight, three in panicle weight, and two for days to heading (**Figure 2.6 & 2.7**). GWA loci were identified from both teff subgenomes with 15 in subgenome A and 11 in subgenome B. Subgenome A has a higher nucleotide diversity and a higher breadth of gene expression, which aligns with the expectation of a higher number of significant loci <sup>24</sup>.

A quarter of the significant loci were found to be pleiotropic, including a SNP on Chromosome 8B (1,305,966 bp) that was identified for both height and lodging GWA and is 1.9 kb from a locus significant for panicle architecture at 1,307,870 bp. This second SNP falls within the coding region of Et\_8B\_058648, an uncharacterized gene. Another set of loci on Chromosome 1A are significant in lodging and height (13,819,711 & 13,821,310 bp) are separated by only 1.6 kb. There is also a significant locus for panicle weight 18kb away on Chromosome 1A (13,837,402 bp). These loci fall between genes Et\_1A\_005766, Et\_1A\_005769, and Et\_1A\_008800. Et\_1A\_005766 shares significant sequence identity with aspartic protease pepsin-like genes that may be associated with grain weight <sup>25,26</sup>. To investigate loci further, genes within 100Kb of the significant SNPs were selected and annotated with GO and KEGG terms. The LD decay ( $r^2$ ) of the panel at 0.1 and 0.2 is 200 and 68.5 Kb respectively, so we were confident 100Kb was sufficient for gene investigation. Orthologous genes in rice, sorghum, and maize were reviewed for putative gene candidates, and protein sequences were annotated for sequence similarity with NCBI protein BLAST. For these pleiotropic loci a few genes were significantly classified into GO terms. Et\_8B\_059059 is a biotin synthase associated with Chromosome 8B (1,305,966 bp) and Chromosome 8B (1,307,870 bp) (**Table 2.5**). For the loci on Chromosome 1A, two genes, Et\_1A\_005778 and Et\_1A\_008802, within the inositol trisphosphate metabolic process were found significant and a barley ortholog was listed as a candidate for lodging resistance <sup>27</sup>. Functional validation of these putative candidate genes is needed in future studies to verify their involvement in lodging tolerance and teff morphology.

#### ***Significant loci associated with lodging***

Seven loci were significantly associated with lodging with p-values less than the Bonferroni threshold. Notably, one major loci on Chromosome 9A (22,621,012 bp) is within

the 5'UTR of gene Et\_9A\_062410. This teff gene is orthologous to the rice gene ESP2 (enclosed shorter panicle 2; LOC\_Os01g02890) which is involved in panicle exertion. The recessive mutant *esp2* produces panicles that are enclosed by flag leaf sheaths and have a shortened uppermost internode without affecting other internode lengths <sup>28,29</sup>. This was further evaluated as SUI1 and SUI3 (shortened uppermost internode 1 and 3) <sup>30-32</sup>. Changes to this gene may have deleterious effects on panicle exertion at the heading stage. We argue a reduced upper internode length lowers the plants center of gravity, thereby reducing lodging tolerance <sup>33,34</sup>. Additional evaluation of the lodging tolerant varieties in this panel is necessary for functional validation. Of interest is also a major locus on Chromosome 4B (12,168,782) that lies within noncoding region of Et\_4B\_036872, a cytochrome P450 76M5 like protein involved in diterpenoid biosynthesis and fungal defense; however, cytochrome P450 enzymes are known to be involved in gibberellic acid bioactivity and have been reported inducing semi-dwarfism in rice <sup>35,36</sup>.

### ***Significant loci associated with panicle architecture***

We also identified a locus significantly associated with panicle architecture within the exon of Et\_2A\_018371 on Chromosome 2A (24,941,867 bp). This gene has high sequence similarity to 3-ketoacyl-CoA synthase which is involved in leaf cuticular wax biosynthesis as well panicle development. The screw flag leaf gene (SFL1) was mapped to a 3-ketoacyl-CoA synthase in rice. The *sfl1* mutant has a screw leaf as well as a screw panicle phenotype in which the panicle branches twist in a screw pattern from the internode <sup>37</sup>. We see a similar screw shape in compact teff panicles. In this study *sfl1* mutants also exhibit dn-type dwarfism in which all internode lengths are reduced. Additional loci for panicle architecture and farmer appreciation were highlighted at Chromosome 2A (24,818,955 bp) 123 Kb away in previous research <sup>38</sup>. Further investigation of this gene region in teff is necessary to determine its effect on both panicle architecture and lodging tolerance <sup>39</sup>.

The Et\_4B\_037033 is 180 Kb from a significant hit on Chromosome 4B (13,698,042). Orthologous genes in rice and sorghum have been mapped to dense and erect panicle (DEP1) and erect panicle (EP) genes. DEP1 is known to reduce the length of the inflorescence internode similar to SUI and increase the number of grains per panicle and thereby yield <sup>40-43</sup>. EP panicles maintain erectness throughout flowering to maturity; they tend to have higher photosynthetic efficiency, thicker stems, and high yields <sup>40-43</sup>. These candidates may serve as crucial markers for teff breeding of high yielding, lodging resistant varieties.

## Discussion

### *Utilizing the Teff Association Panel to evaluate teff phenotypic diversity*

To investigate and quantify the natural genetic variation in teff, a panel of 265 diverse accessions consisting of 259 teff and 6 pilosa were phenotyped for lodging tolerance, panicle architecture, height, days to heading, culm width, panicle weight, and seed per panicle weight. Overall, the panel had high variability across all traits except culm width. By evaluating panicle morphology as well as lodging susceptibility we found that accessions with more open architecture lodged more consistently. Within the panel, subpopulations 4 and 8 stood out as elite germplasm, with taller plants, more compact panicle architecture, and high yield per panicle and may serve as potential parental lines for improved lines with lodging resistance that maintain yield and height.

To evaluate phenotypic variability and isolate loci associated with agronomic traits we used the distribution of natural genetic variation in the TAP to perform GWA. We found loci significantly associated with each trait except for culm width. We aim to optimize the use of the TAP as a resource for the teff breeding community by pairing the genetic resources with phenotypic diversity collected in the field. The development of SSR markers and fine mapping of loci and putative candidate genes would aid in the downstream efforts of precision molecular breeding to enhance lodging tolerance and agronomic improvement.

### *Association between lodging tolerance and plant morphology*

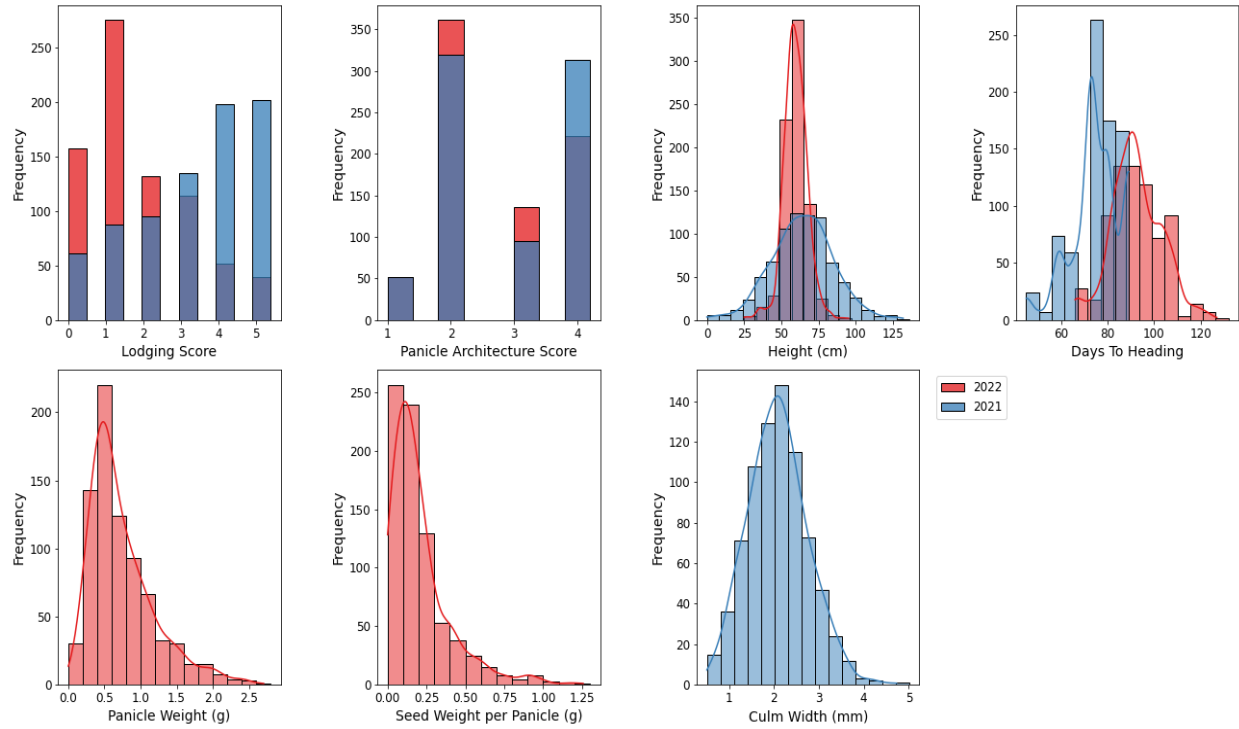
Lodging is a complex trait known to be associated with environmental effects, predominantly wind and rain, as well as a multitude of root, shoot, and panicle morphological traits. During the Green Revolution a negative relationship was identified between lodging resistance and plant height and breeders began selecting for shorter plants with reduced lodging and consistent yield focusing on plant growth hormones: gibberellins, brassinosteroids, and auxins. Two dwarfing genes are the most well characterized; *Rht-1* and *SD-1* which cause insensitivity to gibberellic acid resulting in reduced stem length <sup>44-46</sup>. Recent advancements in teff breeding have developed a semi-dwarf RIL population from a cultivar crossed with an EMS mutant and produced two edited semi-dwarf teff lines with knockout mutation of *SD-1* via *Agrobacterium*-mediated transformation <sup>47,48</sup>. The *SD-1* edited lines showed no change to panicle form and offer an exciting opportunity to increase the lodging tolerance of future knockouts by breeding for panicle architecture traits as well.

Teff fields are consistently lodging across growing regions, and we argue that addressing a combination of traits could rapidly improve teff cultivars. Our results show a negative correlation between lodging severity and height, days to heading, panicle weight, and seed weight as well as positive correlation with panicle architecture score. Although inconsistent with previous studies, the TAP has a high percentage of short lines with lighter, more open panicles as well as a high percentage of tall lines with heavy and more compact panicles. These trends have influenced the association of panicle compactness and lodging tolerance. More compact teff panicles are also more erect and have a smaller panicle angle and thereby lower center of gravity, increasing their lodging tolerance <sup>15</sup>. In addition, plants with erect panicles are less likely to initiate lodging with neighboring plants and are easier to mechanically harvest. The relationship between semi compact panicles and lodging tolerance was supported previously in teff research, but quickly excused with the assumption of a lighter panicle weight <sup>14</sup>. Within the TAP, however, we saw consistent panicle and seed weights across all panicle architectures and lodging severity (**Fig. 2.8**). We argue that the key to maintaining height, yield, plant biomass, and lodging tolerance in teff varieties is an erect, compact panicle. However, the final solution to all crop improvements is adoption, and farmers tend to prefer larger, more open panicles. Additional participatory studies with compact and open teff panicles need to be conducted within Ethiopia and across the globe to determine the primary driver for panicle preference.

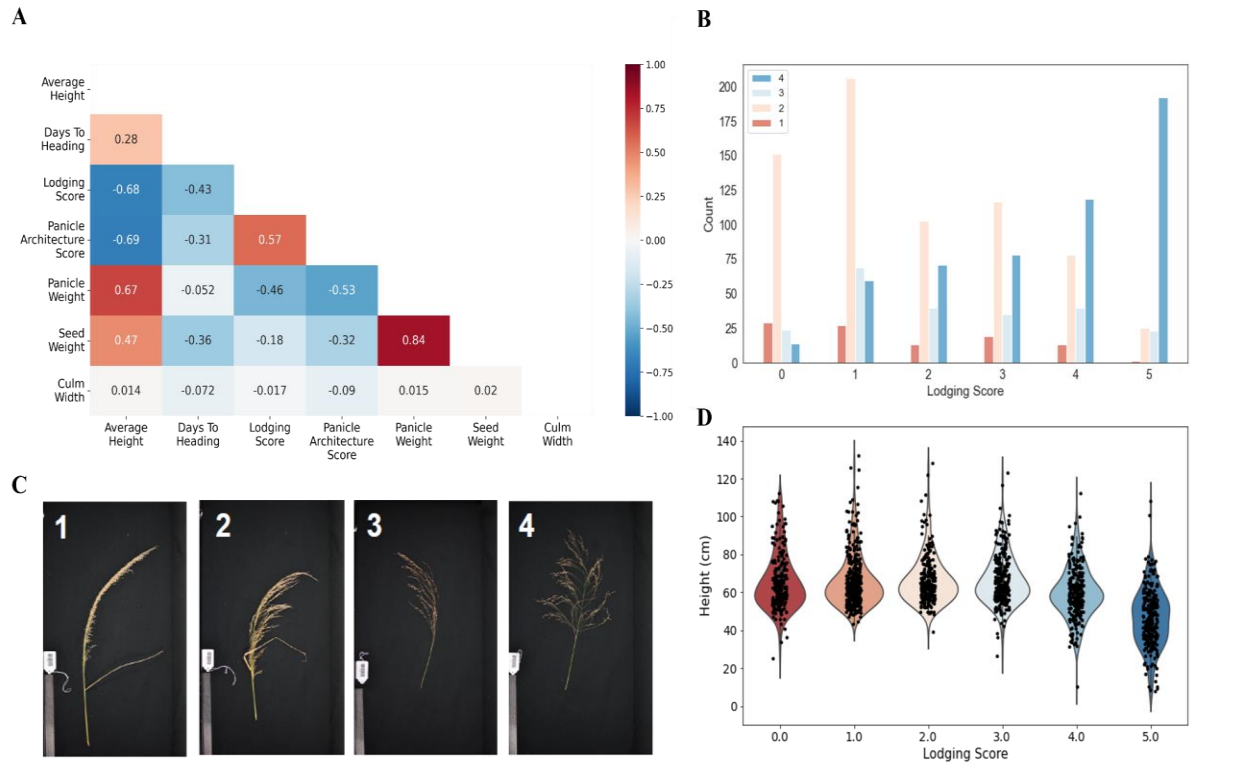
## **Conclusion**

The teff association panel has high genetic and phenotypic diversity that can be harnessed to improve lodging susceptibility and maintain preference across traits. We targeted traits associated with lodging tolerance to provide additional breeding targets outside of dwarfing. We found that panicles with more open architecture lodge more consistently and more compact panicles can improve lodging tolerance while maintaining plant height and panicle yield. 26 loci were identified across 6 agronomic and morphological traits that can be used for marker assisted selection and future breeding efforts.

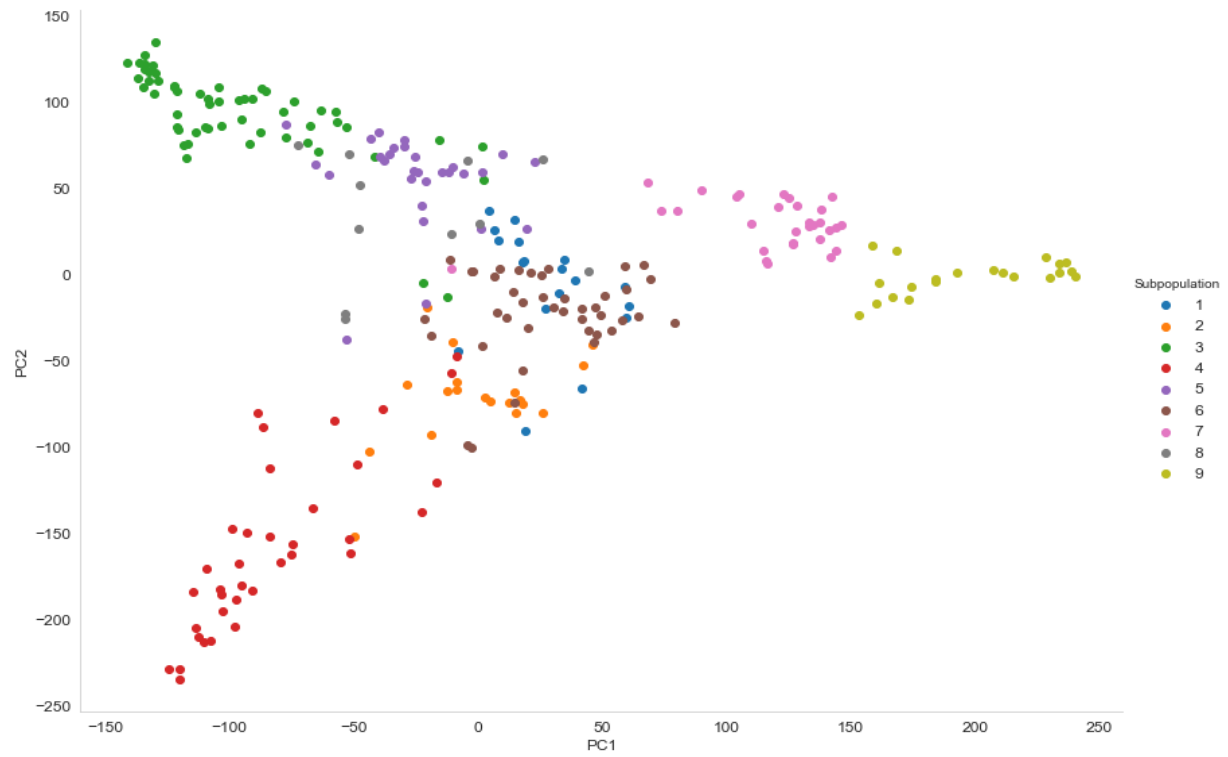
## Figures & Tables



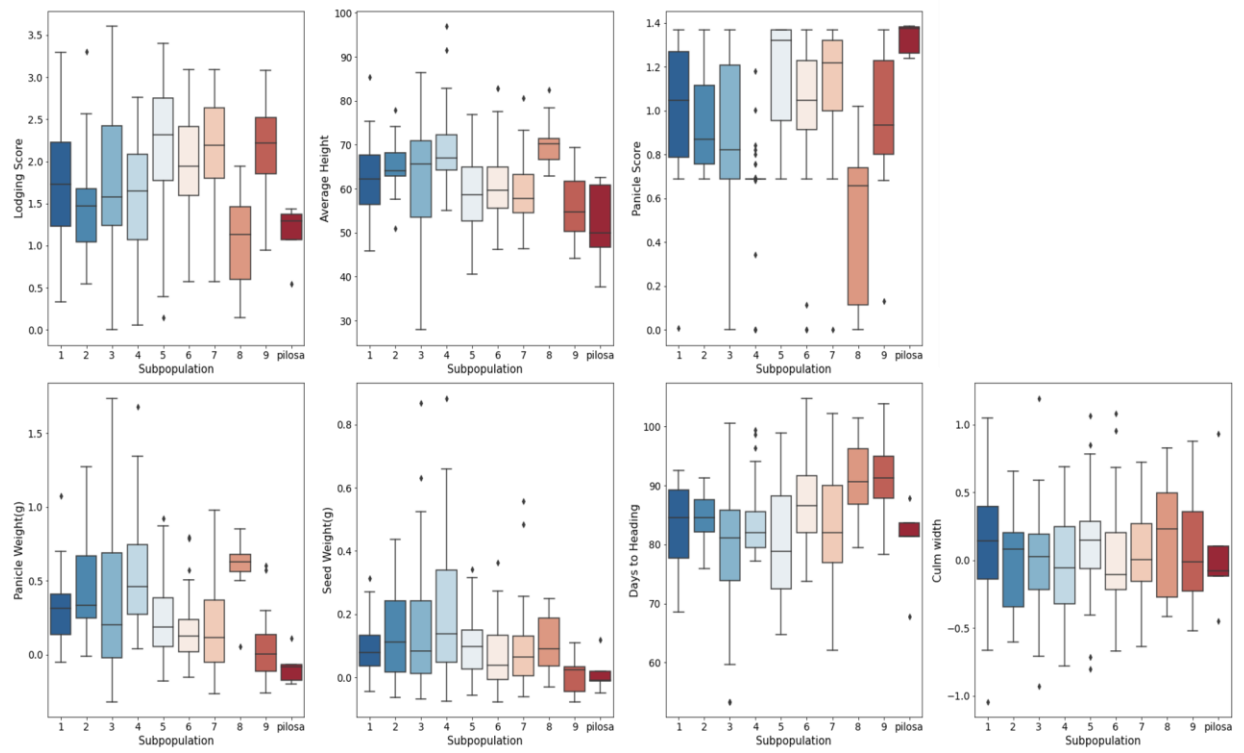
**Figure 2.1: Phenotypic distribution of agronomic and plant architecture in the teff association panel.** Agronomic traits including lodging score, panicle architecture score, height and days to heading are plotted for the TAP from field data in 2020 and 2021. Panicle and seed weight were only collected in 2022, and culm width was only collected in 2021.



**Figure 2.2: Phenotypic associations between developmental traits in teff and lodging susceptibility.** A) Correlation among traits (BLUE values) B) Distribution of panicle architecture by lodging score for both years C) Visual representation of panicle architecture scores D) Distribution of plant height by lodging score for both years.

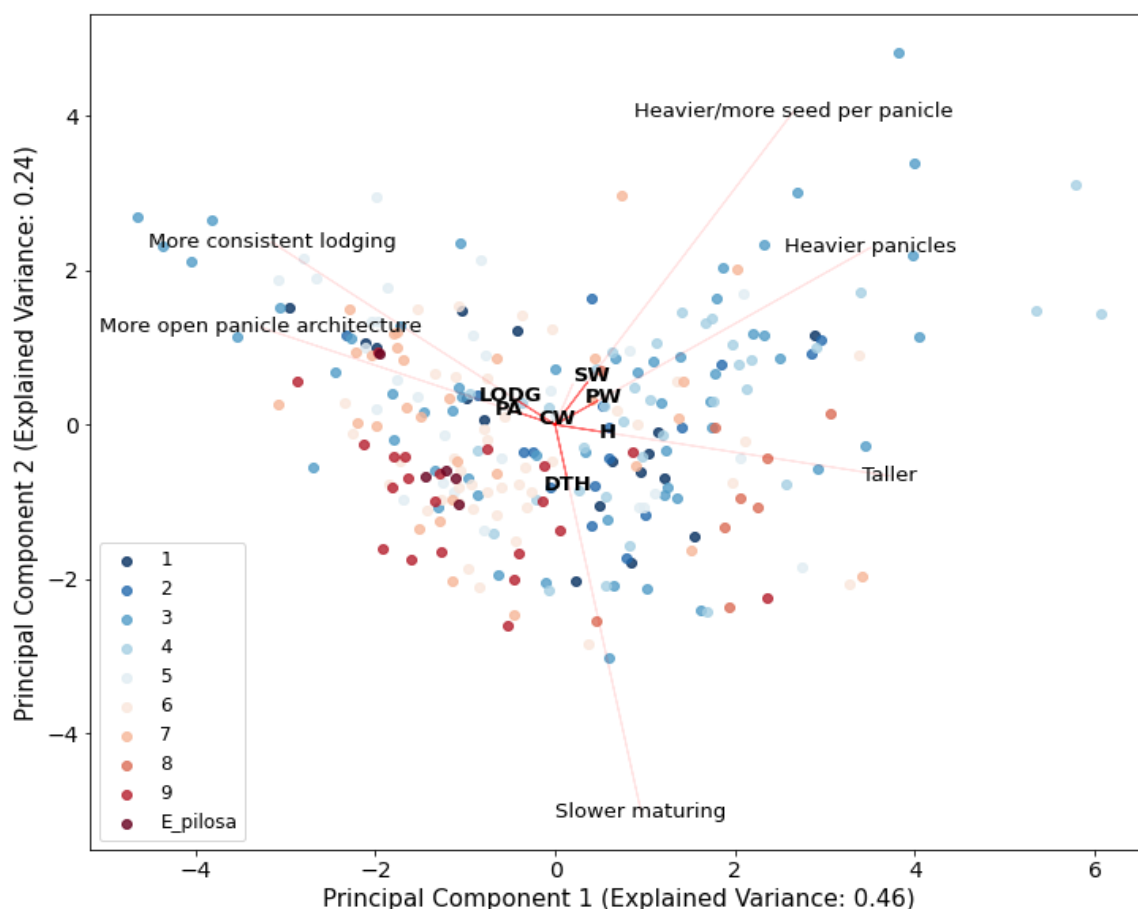


**Figure 2.3: Principal component analysis of genetic variation and population structure across the panel.**

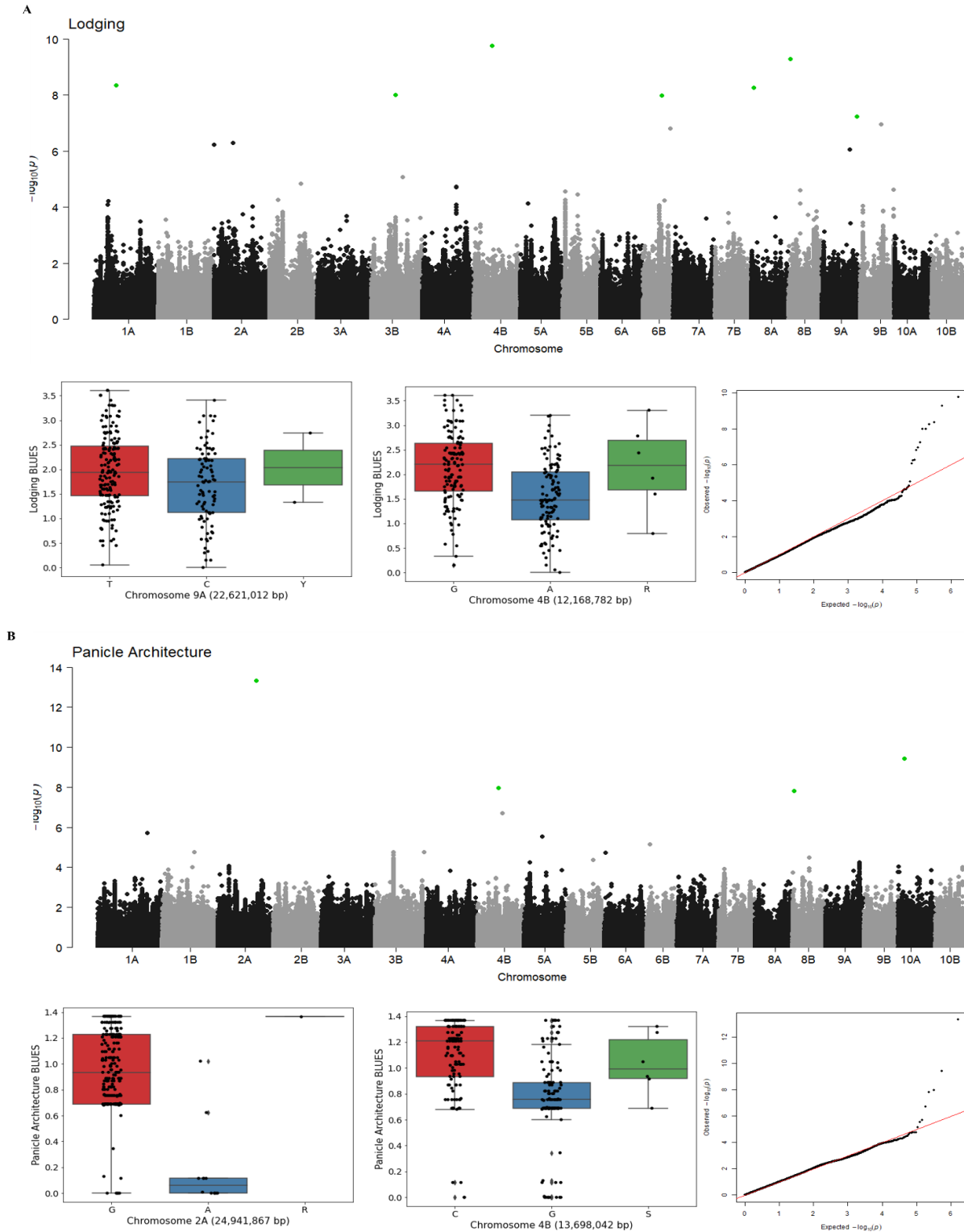


**Figure 2.4: Phenotypic variation (BLUES) across teff subpopulations and *Eragrostis pilosa* accessions.**

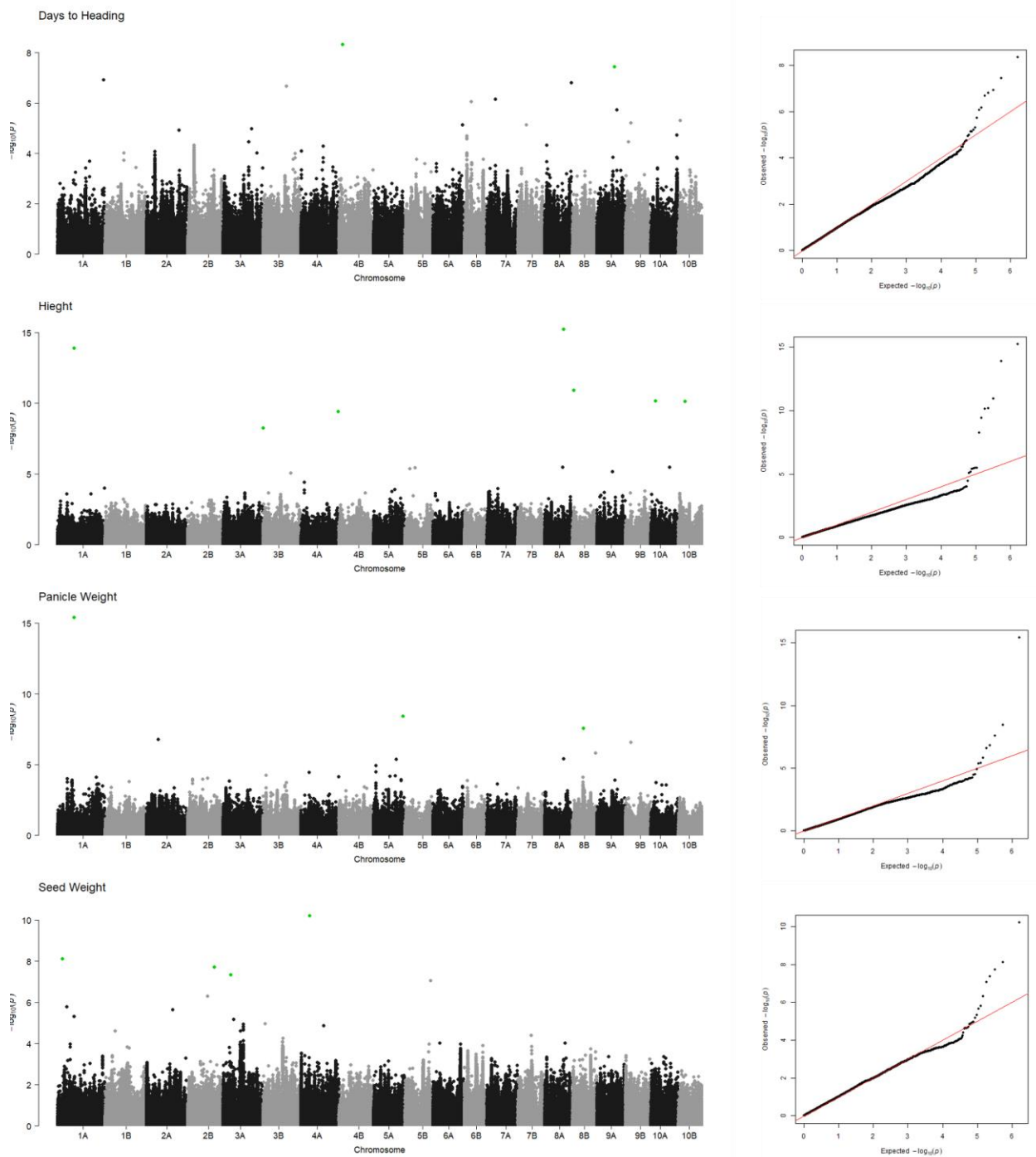




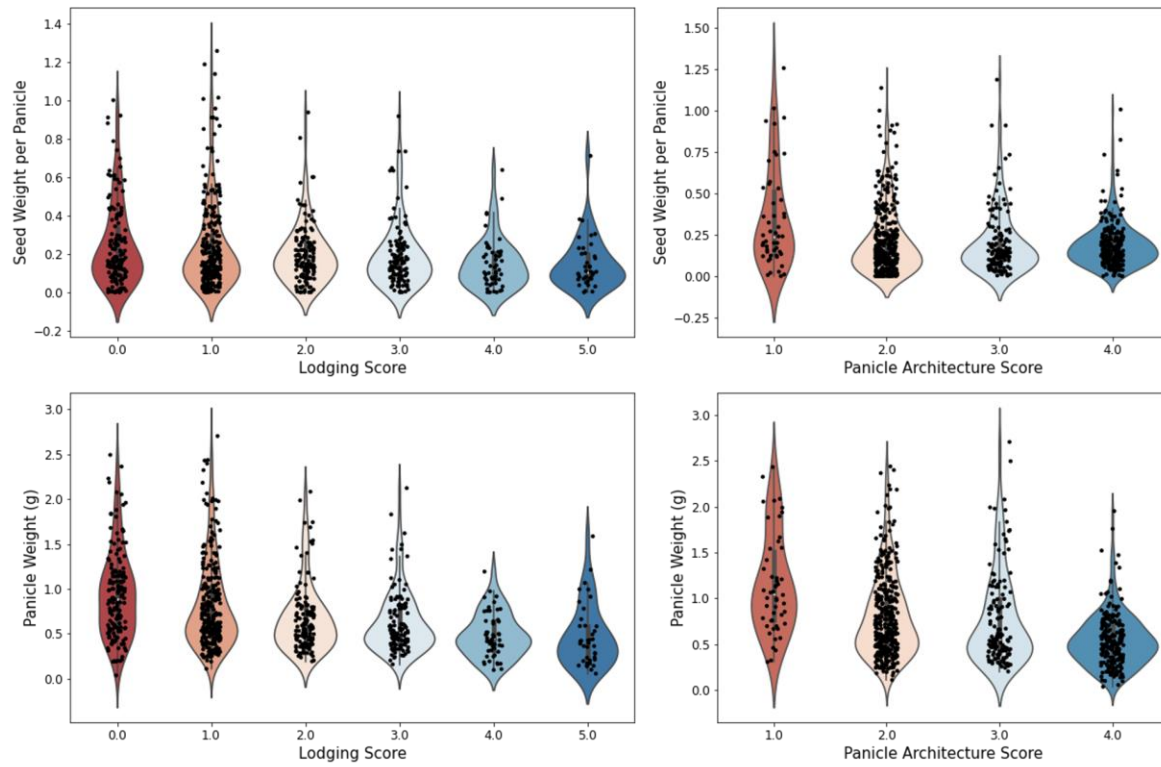
**Figure 2.5: Trait biplot of BLUES, colored by subpopulation.** Explanation of vector abbreviations and trends: (LODG) higher lodging score, more consistent lodging, (PA) panicle architecture score, more open panicle architecture, (DTH) days to heading, slower maturing, (H) Height, taller, (PW) three panicle weight, heavier panicles, (SW) seed weight of three panicles, heavier/more seed per panicle.



**Figure 2.6: FarmCPU genome wide association manhattan plots.** Significant loci highlighted in green, allele frequency of loci discussed and QQ plots in A) Lodging score B) Panicle Architecture.



**Figure 2.7: FarmCPU genome wide association manhattan and QQ plots for days to heading, height, panicle weight, and seed weight, significant loci highlighted in green.**



**Figure 2.8: Panicle and seed weight distribution across lodging and panicle architecture score.**

**Table 2.1:** Phenotypic diversity metrics.

<b>Trait</b>	<b>Year</b>	<b>Mean <math>\pm</math> Sd</b>	<b>Coefficient Of Variation (%)</b>	<b><math>h^2a</math></b>	<b><math>\sigma^2g</math></b>	<b><math>\sigma^2r</math></b>
Lodging Severity	2021- 2022	2.44 $\pm$ 1.68	68.93	0.735587644	0.479165486	0.688958147
Panicle Architecture	2021- 2022	2.77 $\pm$ 1.0	36.02	0.927212258	0.115388079	0.036232643
Height (cm)	2021- 2022	61.83 $\pm$ 15.84	25.63	0.630948417	73.68057498	172.387676
Days to Heading	2021- 2022	83.34 $\pm$ 13.82	16.58	0.820223284	64.46641719	56.51905293
Panicle weight (g)	2022	0.74 $\pm$ 0.45	61.03	0.836861438	0.128386234	0.075083202
Seed weight (g)	2022	0.21 $\pm$ 0.20	94.97	0.773624475	0.019822447	0.017401144
Culm width (mm)	2021	2.09 $\pm$ 0.67	31.87	0.148765556	0.022221184	0.381447922

**Table 2.2:** ANOVA results from linear models across agronomic traits, significance codes: 0 \*\*\*0.001, \*\*0.01, \*0.05.

<b>Trait</b>	<b>Accession Pvalue</b>	<b>Subpopulation Pvalue</b>	<b>Block Pvalue</b>	<b>Year Pvalue</b>
Lodging Severity	<2.2e-16 ***	<2.2e-16 ***	3.133e-09 ***	<2.2e-16 ***
Panicle Architecture	<2.2e-16 ***	<2.2e-16 ***	0.08445	5.831e-09 ***
Height (cm)	<2.2e-16 ***	<2.2e-16 ***	0.5202	3.653e-12 ***
Days to Heading	< 2.2e-16 ***	< 2.2e-16 ***	0.04569 *	< 2.2e-16 ***
Panicle weight (g)	< 2.2e-16 ***	< 2.2e-16 ***	0.004896 **	
Seed weight (g)	< 2.2e-16 ***	< 2.2e-16 ***	2.703e-13 ***	
Culm width (mm)	0.06443	0.52651	< 2e-16 ***	

**Table 2.3:** Coefficients of variation across subpopulations for each trait.

Sub pop	Height	Days to Heading	Lodging	Panicle Architecture	Panicle Weight	Seed Weight	Culm Width
1	15.051419	8.650197	48.281986	34.739453	92.457044	105.681100	414.440722
2	9.638622	5.013176	44.519325	24.364396	77.587650	101.675401	15422.714401
3	20.503377	13.286679	49.142672	40.391836	142.354472	128.190579	-2379.976923
4	12.671501	7.134070	41.339907	35.503319	74.099460	102.935771	-1694.204966
5	15.272576	12.068080	37.398675	22.447214	114.067689	100.015587	377.833140
6	13.405369	7.886023	26.682776	33.363005	145.810772	140.666904	-14222.494070
7	13.314261	10.951848	33.215566	27.141773	163.973464	149.768636	712.388194
8	8.799326	8.399625	55.005730	75.842842	38.443657	85.769317	284.918593
9	12.311889	7.516318	27.051331	29.382355	585.536054	1885.206049	618.347880
E_p							
ilos	20.060236	9.348523	31.758352	5.365758	-	473.440866	669.052059
a					145.501417		

**Table 2.4:** Significant GWA loci.

SNP	Chromosome	Position	Pvalue	MAF	Trait
Chromosome_10A_3991573	19	3991573	3.77E-10	0.071705	Panicle Architecture
Chromosome_10A_4046309	19	4046309	7.03E-11	0.054264	Height
Chromosome_10B_5678730	20	5678730	7.34E-11	0.065891	Height
Chromosome_1A_13819711	1	13819711	4.45E-09	0.282946	Lodging
Chromosome_1A_13821310	1	13821310	1.30E-14	0.277132	Height
Chromosome_1A_13837402	1	13837402	3.92E-16	0.267578	Panicle Weight
Chromosome_1A_3973096	1	3973096	7.54E-09	0.054688	Seed Weight
Chromosome_2A_24941867	3	24941867	4.68E-14	0.04845	Panicle Architecture
Chromosome_2B_22941475	4	22941475	1.88E-08	0.3125	Seed Weight
Chromosome_3A_6485330	5	6485330	4.44E-08	0.060547	Seed Weight
Chromosome_3B_15446298	6	15446298	1.02E-08	0.372093	Lodging
Chromosome_3B_68136	6	68136	5.85E-09	0.203488	Height
Chromosome_4A_31876960	7	31876960	3.75E-10	0.162791	Height
Chromosome_4A_7415683	7	7415683	6.24E-11	0.054688	Seed Weight



**Table 2.4 (cont'd)**

Chromosome_4B_12168782	8	12168782	1.72E-10	0.46124	Lodging
Chromosome_4B_13698042	8	13698042	1.07E-08	0.48062	Panicle Architecture
Chromosome_4B_3188738	8	3188738	4.52E-09	0.116279	Days to Heading
Chromosome_5A_25320321	9	25320321	3.80E-09	0.4375	Panicle Weight
Chromosome_6B_12170310	12	12170310	1.05E-08	0.124031	Lodging
Chromosome_8A_16447012	15	16447012	6.08E-16	0.063953	Height
Chromosome_8A_2118480	15	2118480	5.50E-09	0.124031	Lodging
Chromosome_8B_1305966	16	1305966	1.20E-11	0.267442	Height
Chromosome_8B_1305966	16	1305966	5.19E-10	0.267442	Lodging
Chromosome_8B_1307870	16	1307870	1.50E-08	0.292636	Panicle Architecture
Chromosome_8B_9369935	16	9369935	2.70E-08	0.195313	Panicle Weight
Chromosome_9A_15292040	17	15292040	3.60E-08	0.056202	Days to Heading
Chromosome_9A_22621012	17	22621012	5.76E-08	0.352713	Lodging

**Table 2.5:** Significant GO term associations with  $pval < 0.05$  for genes  $\pm 100Kb$  from significant loci.

GO ID	Term	P value	Gene	Chrom	SNP	Gene Start	Gene Stop	Trait
GO: 0007018	microtubule-based movement	0.0582	Et_10A_001494	9	Chromosome 10A_3991573	4060916	4065050	Panicle Architecture
GO: 0007018	microtubule-based movement	0.0582	Et_10A_001494	9	Chromosome 10A_4046309	4060916	4065050	Height
GO: 0007018	microtubule-based movement	0.0582	Et_10A_001499	9	Chromosome 10A_4046309	4102790	4115492	Height
GO: 0002098	tRNA wobble uridine modification	0.0154	Et_10A_001502	9	Chromosome 10A_4046309	4126053	4130830	Height
GO: 0007165	signal transduction	0.0179	Et_10A_002191	9	Chromosome 10A_3991573	3930890	3935153	Panicle Architecture
GO: 0002957	inositol trisphosphate metabolic process	0.0031	Et_1A_005778	1	Chromosome 1A_13819711	13919213	13922418	Lodging

**Table 2.5 (cont'd)**

GO: 0032957	inositol trisphosphate metabolic process	0.0031	Et_1A_005778	1	Chromosome 1A_138213	1391921	1392241	Height Weight
GO: 0032957	inositol trisphosphate metabolic process	0.0031	Et_1A_005778	1	Chromosome 1A_138374	1391921	1392241	Panicle Weight
GO: 0007165	signal transduction	0.0179	Et_1A_007860	1	Chromosome 1A_397309	3934125	3936043	Seed Weight
GO: 0032957	inositol trisphosphate metabolic process	0.0031	Et_1A_008802	1	Chromosome 1A_138197	1391807	1391912	Lodging
GO: 0032957	inositol trisphosphate metabolic process	0.0031	Et_1A_008802	1	Chromosome 1A_138374	1391807	1391912	Panicle Weight
GO: 0007018	microtubule-based movement	0.0582	Et_1A_009324	1	Chromosome 1A_397309	3903124	3910510	Seed Weight
GO: 0006629	lipid metabolic process	0.0206	Et_2A_018371	3	Chromosome 2A_249418	2494054	2494198	Panicle Architecture

**Table 2.5 (cont'd)**

GO: 0018342	protein prenylation	0.0382	Et_2B_020698	4	Chromosome _2B_229414	2288470	2288689	Seed Weight
						4	3	75
GO: 0006629	lipid metabolic process	0.0206	Et_2B_020712	4	Chromosome _2B_229414	2298025	2298419	Seed Weight
						3	7	75
GO: 0006629	lipid metabolic process	0.0206	Et_2B_020712	4	Chromosome _2B_229414	2298025	2298419	Seed Weight
						3	7	75
GO: 0018342	protein prenylation	0.0382	Et_3A_026289	5	Chromosome _3A_648533	6466457	6476247	Seed Weight
								0
GO: 0006629	lipid metabolic process	0.0206	Et_3A_026290	5	Chromosome _3A_648533	6461665	6466335	Seed Weight
								0
GO: 0006376	mRNA splice site recognition	0.0306	Et_4A_034332	7	Chromosome _4A_318769	3184699	3185183	Height
						0	7	60
GO: 0007165	signal transduction	0.0179	Et_4A_034336	7	Chromosome _4A_318769	3188702	3189125	Height
						4	9	60

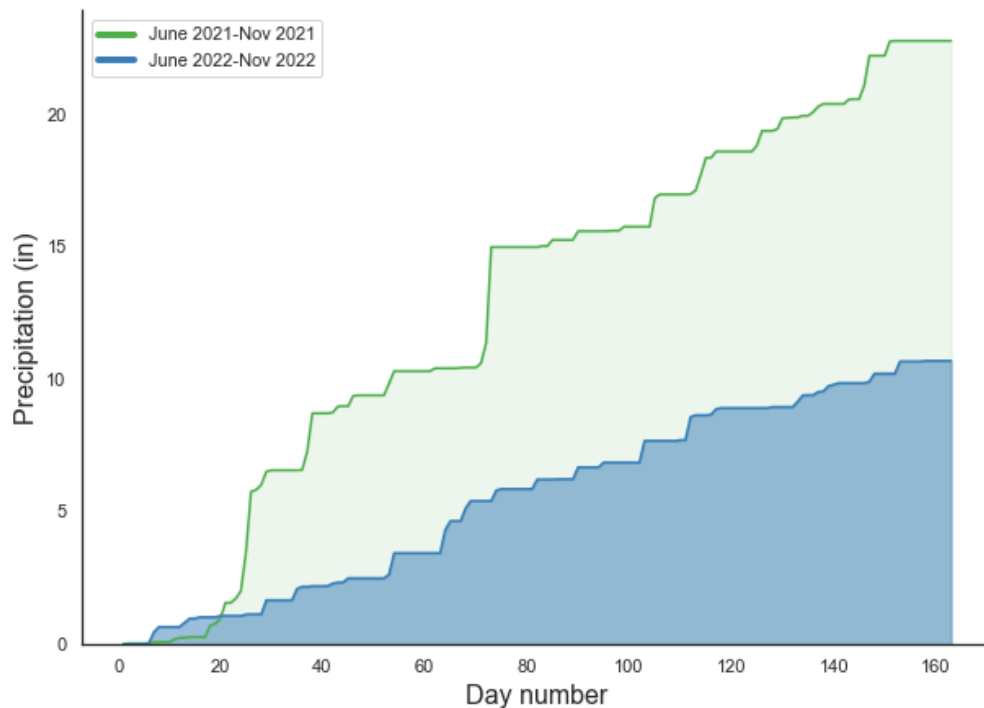
**Table 2.5 (cont'd)**

GO: lipid	0.0206	Et_4B_	8	Chromosome	1221360	1222009	Lodging
000 metabolic		036876		_4B_121687	9	4	
662 process				82			
9							
GO: steroid	0.0113	Et_4B_	8	Chromosome	1377823	1379477	Panicle
000 biosynthet		037027		_4B_136980	7	2	Architecture
669 ic process				42			
4							
GO: steroid	0.0113	Et_4B_	8	Chromosome	1377823	1379477	Panicle
000 biosynthet		037027		_4B_136980	7	2	Architecture
669 ic process				42			
4							
GO: biotin	0.0154	Et_8B_	16	Chromosome	1333142	1344961	Height &
000 biosynthet		059053		_8B_130596			Lodging
910 ic process				6			
2							
GO: biotin	0.0154	Et_8B_	16	Chromosome	1333142	1344961	Panicle
000 biosynthet		059053		_8B_130787			Architecture
910 ic process				0			
2							
GO: biotin	0.0154	Et_8B_	16	Chromosome	1325006	1328927	Height &
000 biosynthet		059059		_8B_130596			Lodging
910 ic process				6			
2							
GO: biotin	0.0154	Et_8B_	16	Chromosome	1325006	1328927	Panicle
000 biosynthet		059059		_8B_130787			Architecture
910 ic process				0			
2							
GO: lipid	0.0206	Et_9A_	17	Chromosome	2254566	2254953	Lodging
000 metabolic		062404		_9A_226210	8	3	
662 process				12			
9							

**Table 2.5 (cont'd)**

GO: 0006537	glutamate biosynthetic process	0.0456	Et_9A_062407	17	Chromosome_9A_226210	22576546	22586893	Lodging
GO: 0007165	signal transduction	0.0179	Et_9A_062409	17	Chromosome_9A_226210	22599487	22601465	Lodging

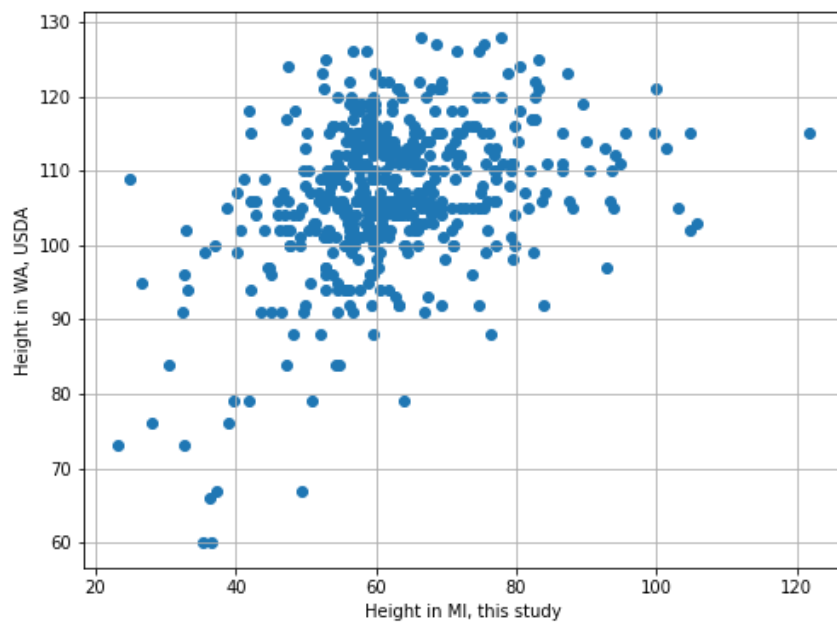
## Supplemental



**Figure Supplemental 2.1: Cumulative precipitation June 1st through November 10th in 2021 and 2022 from the weather station at East Lansing (HTRC).**

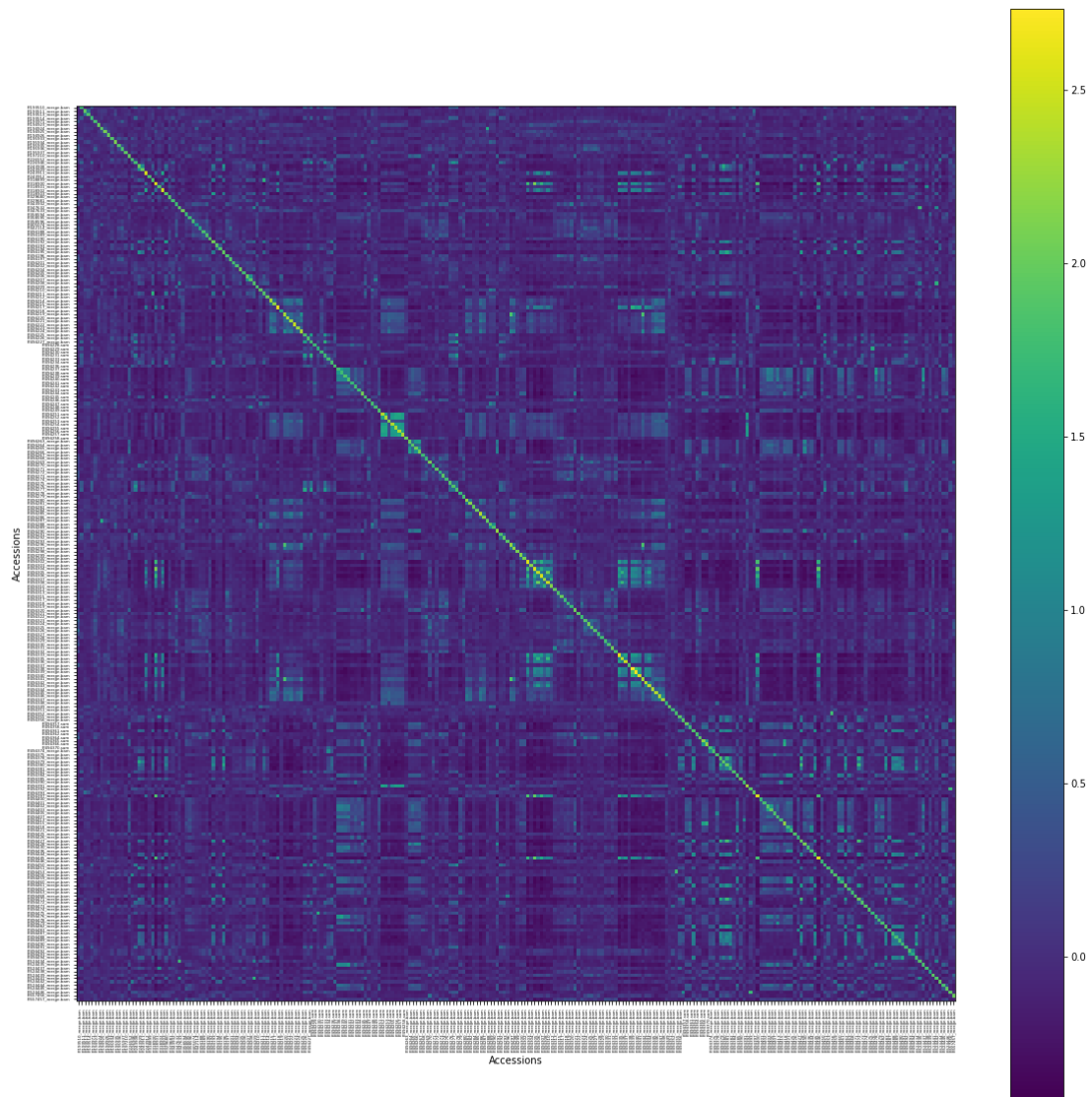
***Supplemental methods:***

Temperature, rainfall, and degree-day summary were collected for the growing season in 2021 and 2022, June 1st through November 10th from the East Lansing, Michigan Horticulture Teaching and Research Center weather station at the Hancock Turgrass Research Center (42.7110, -84.4760) (<https://legacy.enviroweather.msu.edu/weather.php?stn=htc>).



**Figure Supplemental 2.2: Comparison of height data collected in our field experiment to those on the USDA GRIN website.**





**Figure Supplemental 2.3: Kinship matrix of TAP.**

**Table Supplemental 2.1:** Results from probit ordinal regression, lodging score was increased by 1 to eliminate numeric data of 0.

**2021**

<b>Dep. Variable:</b>	LodgCategory	<b>Log-Likelihood:</b>	-1010.7			
<b>Model:</b>	OrderedModel	<b>AIC:</b>	2039.			
<b>Method:</b>	Maximum Likelihood	<b>BIC:</b>	2081.			
<b>No. Observations:</b>	761					
<b>Df Residuals:</b>	752					
<b>Df Model:</b>	9					
	<b>coef</b>	<b>std err</b>	<b>z</b>	<b>P&gt; z </b>	<b>[0.025</b>	<b>0.975]</b>
<b>Panicle Architecture</b>	0.3698	0.048	7.687	0.000	0.275	0.464
<b>Height</b>	-0.0340	0.003	-12.409	0.000	-0.039	-0.029
<b>Days to Heading</b>	-0.0403	0.004	-9.517	0.000	-0.049	-0.032
<b>Culm Width</b>	-0.0396	0.070	-0.565	0.572	-0.177	0.098
<b>1.0/2.0</b>	-6.4006	0.457	-14.008	0.000	-7.296	-5.505
<b>2.0/3.0</b>	-0.2808	0.100	-2.820	0.005	-0.476	-0.086
<b>3.0/4.0</b>	-0.5678	0.094	-6.008	0.000	-0.753	-0.383
<b>4.0/5.0</b>	-0.3653	0.078	-4.679	0.000	-0.518	-0.212
<b>5.0/6.0</b>	0.1188	0.063	1.891	0.059	-0.004	0.242

**Table Supplemental 2.1 (cont'd)**

**2022**

Dep. Variable:		LodgCategory		Log-Likelihood:		-978.28	
Model:		OrderedModel		AIC:		1977.	
Method:		Maximum Likelihood		BIC:		2022.	
No. Observations:		688					
Df Residuals:		678					
Df Model:		10					
		coef	std err	z	P> z	[0.025	0.975]
Panicle Architecture		0.3644	0.052	7.064	0.000	0.263	0.465
Height		0.0207	0.007	2.911	0.004	0.007	0.035
Days to Heading		-0.0213	0.005	-4.151	0.000	-0.031	-0.011
Panicle Weight		-1.1063	0.185	-5.982	0.000	-1.469	-0.744
Seed Weight		0.4756	0.408	1.166	0.243	-0.324	1.275
1.0/2.0		-1.4897	0.786	-1.895	0.058	-3.031	0.051
2.0/3.0		0.2083	0.054	3.853	0.000	0.102	0.314
3.0/4.0		-0.5411	0.083	-6.501	0.000	-0.704	-0.378
4.0/5.0		-0.3622	0.090	-4.010	0.000	-0.539	-0.185
5.0/6.0		-0.5131	0.138	-3.705	0.000	-0.784	-0.242

## REFERENCES

1. Lee, H. Teff, A Rising Global Crop: Current Status of Teff Production and Value Chain. *The Open Agriculture Journal* vol. 12 185–193 Preprint at <https://doi.org/10.2174/1874331501812010185> (2018).
2. Saylor, B. A., Min, D. H. & Bradford, B. J. Productivity of lactating dairy cows fed diets with teff hay as the sole forage. *Journal of Dairy Science* vol. 101 5984–5990 Preprint at <https://doi.org/10.3168/jds.2017-14118> (2018).
3. Barretto, R. *et al.* Teff ( *Eragrostis tef* ) processing, utilization and future opportunities: a review. *International Journal of Food Science & Technology* vol. 56 3125–3137 Preprint at <https://doi.org/10.1111/ijfs.14872> (2021).
4. D’Andrea, A. C. T’ef (*Eragrostis tef*) in Ancient Agricultural Systems of Highland Ethiopia. *Econ. Bot.* **62**, 547–566 (2008).
5. Woldeyohannes, A. B. *et al.* Data-driven, participatory characterization of farmer varieties discloses teff breeding potential under current and future climates. Preprint at <https://doi.org/10.1101/2021.08.27.457623>.
6. Assefa, K. *et al.* Breeding tef [*Eragrostis tef* (Zucc.) trotter]: conventional and molecular approaches. *Plant Breed.* **130**, 1–9 (2011).
7. Ben-Zeev, S. *et al.* Unraveling the central role of root morphology and anatomy in lodging of tef ( *Eragrostis tef* ). *Plants People Planet* (2023) doi:10.1002/ppp3.10389.
8. Numan, M. *et al.* From Traditional Breeding to Genome Editing for Boosting Productivity of the Ancient Grain Tef [ (Zucc.) Trotter]. *Plants* **10**, (2021).
9. Alemu, M. D. *et al.* Genomic dissection of productivity, lodging, and morpho-physiological traits in *Eragrostis tef* under contrasting water availabilities. *Plants People Planet* (2024) doi:10.1002/ppp3.10505.
10. van Delden, S. H., Vos, J., Ennos, A. R. & Stomph, T. J. Analysing lodging of the panicle bearing cereal teff (*Eragrostis tef*). *New Phytol.* **186**, (2010).
11. Ben-Zeev, S. *et al.* Less Is More: Lower Sowing Rate of Irrigated Tef (*Eragrostis tef*) Alters Plant Morphology and Reduces Lodging. *Agronomy* vol. 10 570 Preprint at <https://doi.org/10.3390/agronomy10040570> (2020).
12. Jifar, H. *et al.* Semi-dwarf tef lines for high seed yield and lodging tolerance in Central Ethiopia. *Afr. Crop Sci. J.* **25**, 419 (2017).
13. Beyene, G. *et al.* CRISPR/Cas9-mediated tetra-allelic mutation of the ‘Green Revolution’ SEMIDWARF-1 (SD-1) gene confers lodging resistance in tef (*Eragrostis tef*). *Plant Biotechnol. J.* **20**, 1716–1729 (2022).

14. Bayable, M. *et al.* Biomechanical Properties and Agro-Morphological Traits for Improved Lodging Resistance in Ethiopian Teff (*Eragrostis tef* (Zucc.) Trotter) Accessions. *Agronomy* **10**, 1012 (2020).
15. Blösch, R. *et al.* Panicle Angle is an Important Factor in Tef Lodging Tolerance. *Front. Plant Sci.* **11**, 61 (2020).
16. Ben-Zeev, S. *et al.* Less Is More: Lower Sowing Rate of Irrigated Tef (*Eragrostis tef*) Alters Plant Morphology and Reduces Lodging. *Agronomy* vol. 10 570 Preprint at <https://doi.org/10.3390/agronomy10040570> (2020).
17. Assefa, K. *et al.* Genetic diversity in tef [*Eragrostis tef* (Zucc.) Trotter]. *Front. Plant Sci.* **6**, 177 (2015).
18. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S*. (Springer Science & Business Media, 2013).
19. statsmodels 0.15.0 (+263). <https://www.statsmodels.org/devel/>.
20. Website. The pandas development team. (2024). pandas-dev/pandas: Pandas (v2.2.2). Zenodo. <https://doi.org/10.5281/zenodo.10957263>.
21. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
22. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics Proteomics Bioinformatics* **19**, 629–640 (2021).
23. D. Turner, S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
24. VanBuren, R. *et al.* Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
25. Niu, K.-X. *et al.* Suppressing ASPARTIC PROTEASE 1 prolongs photosynthesis and increases wheat grain weight. *Nature Plants* **9**, 965–977 (2023).
26. Genome-wide association mapping for seed protein content in finger millet (*Eleusine coracana*) global collection through genotyping by sequencing. *J. Cereal Sci.* **91**, 102888 (2020).
27. Stem lodging resistance in hulless barley: Transcriptome and metabolome analysis of lignin biosynthesis pathways in contrasting genotypes. *Genomics* **113**, 935–943 (2021).
28. Dang, X. *et al.* A new allele PEL9GG identified by genome-wide association study increases panicle elongation length in rice (*Oryza sativa* L.). *Front. Plant Sci.* **14**, 1136549 (2023).

29. Guan, H. *et al.* Genetic analysis and fine mapping of an enclosed panicle mutant *esp2* in rice (*Oryza sativa* L.). *Chin. Sci. Bull.* **56**, 1476–1480 (2011).
30. Zhu, L. *et al.* Identification and characterization of SHORTENED UPPERMOST INTERNODE 1, a gene negatively regulating uppermost internode elongation in rice. *Plant Mol. Biol.* **77**, 475–487 (2011).
31. Yin, H. *et al.* SUI-family genes encode phosphatidylserine synthases and regulate stem development in rice. *Planta* **237**, 15–27 (2012).
32. Rani, M. H. *et al.* ES5 is involved in the regulation of phosphatidylserine synthesis and impacts on early senescence in rice (*Oryza sativa* L.). *Plant Mol. Biol.* **102**, 501–515 (2020).
33. Zhu, Y. *et al.* ELONGATED UPPERMOST INTERNODE Encodes a Cytochrome P450 Monooxygenase That Epoxidizes Gibberellins in a Novel Deactivation Reaction in Rice. *Plant Cell* **18**, 442–456 (2006).
34. Liu, C. *et al.* Shortened Basal Internodes Encodes a Gibberellin 2-Oxidase and Contributes to Lodging Resistance in Rice. *Mol. Plant* **11**, 288–299 (2018).
35. Niu, Y., Chen, T., Zhao, C. & Zhou, M. Improving Crop Lodging Resistance by Adjusting Plant Height and Stem Strength. *Agronomy* **11**, 2421 (2021).
36. Zhao, D.-D., Son, J.-H., Farooq, M. & Kim, K.-M. Identification of Candidate Gene for Internode Length in Rice to Enhance Resistance to Lodging Using QTL Analysis. *Plants* **10**, (2021).
37. Alamin, M. *et al.* Characterization and Fine Mapping of SFL1, a Gene Controlling Screw Flag Leaf in Rice. *Plant Mol. Biol. Rep.* **35**, 491–503 (2017).
38. Woldeyohannes, A. B. *et al.* Data-driven, participatory characterization of farmer varieties discloses teff breeding potential under current and future climates. (2022) doi:10.7554/eLife.80009.
39. Cheng, X. *et al.* Potentially Useful Dwarfing or Semi-dwarfing Genes in Rice Breeding in Addition to the *sd1* Gene. *Rice* **15**, (2022).
40. Huang, X. *et al.* Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* **41**, 494–497 (2009).
41. Wang, J. *et al.* Identification and characterization of the erect-pose panicle gene EP conferring high grain yield in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **119**, 85–91 (2009).
42. Zhou, Y. *et al.* Deletion in a Quantitative Trait Gene qPE9-1 Associated With Panicle Erectness Improves Plant Architecture During Rice Domestication. *Genetics* **183**, 315 (2009).
43. Tao, Y. *et al.* Manipulating assimilate availability provides insight into the genes

- controlling grain size in sorghum. *Plant J.* **108**, 231–243 (2021).
44. Gale, M. D. & Marshall, G. A. Insensitivity to gibberellin in dwarf wheats. *Ann. Bot.* **37**, 729–735 (1973).
  45. Monna, L. *et al.* Positional cloning of rice semidwarfing gene, sd-1: rice ‘green revolution gene’ encodes a mutant enzyme involved in gibberellin synthesis. *DNA Res.* **9**, (2002).
  46. Sasaki, A. *et al.* Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* **416**, (2002).
  47. Tadesse, M. *et al.* Evaluation of selected semi-dwarf Tef (*Eragrostis tef* (Zucc.) Trotter) genotypes for yield and yield related traits. *J. Sci. Agric.* 6–12 (2022).
  48. Beyene, G. *et al.* CRISPR/Cas9-mediated tetra-allelic mutation of the ‘Green Revolution’ SEMIDWARF-1 (SD-1) gene confers lodging resistance in tef (*Eragrostis tef*). *Plant Biotechnol. J.* **20**, 1716–1729 (2022).

## **CHAPTER 4: GENETIC INSIGHTS AND EXAMINATION OF SEED NUTRIENT VARIABILITY IN THE TEFF ASSOCIATION PANEL**



## Abstract

Teff is a resilient and nutrient-rich cereal with tremendous potential in the agricultural industry. To investigate the seed nutrient variability in teff, we phenotyped a diverse set of 265 teff cultivars, farmer varieties, and the wild progenitor *Eragrostis pilosa* within the Teff Association Panel (TAP) for 12 seed nutrients. We found that Phosphorus (P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sulfur (S), Zinc (Zn), Copper (Cu), and Boron (B) exhibit relatively low variation across the panel, whereas Manganese (Mn), Iron (Fe), and Aluminum (Al) show high variability and abundance. The average Fe content was 46 mg per 100 g, which is considerably higher than other cereals, and validates previously contentious findings. We observed that white seeded accessions had higher concentrations of Zn, Cu, P, K, Mg, and S, while brown seeded accessions had elevated levels of Mg, Al, Fe, and Ca. A genome wide association study across all nutrients identified 19 loci significantly associated with concentrations of P, K, Zn, Fe, and Al. Four of these loci explain greater than 30% of the phenotypic variation for specific nutrient content. The genomic regions associated with these hits can be used for marker assisted selection to enhance the nutritional value of teff.

## Introduction

Leading cereals such as maize, rice, and wheat were selected for optimized yield under intensive cultivation, and stress tolerance was lost as a byproduct of domestication <sup>1</sup>. As a result, leading cereals are susceptible to abiotic stresses, resulting in billions in losses each year <sup>2</sup>. Teff is an exception, and although generally lower yielding than leading cereals, it is adapted to dry conditions unsuitable for other grain crops, making it an attractive alternative grain for production in low rainfall areas. Leading cereals have inherently low micronutrient concentrations, and over-reliance on these crops leads to micronutrient deficiencies. Malnutrition is a global issue that is perpetually evolving, and the number of people affected by hunger has increased 20% since the Covid-19 pandemic <sup>3</sup>. When compared to other cereals like sorghum and maize, teff has a higher nutritional value, including up to three fold as much Ca, Zn, Cu, and Fe <sup>4</sup>. Teff seed is also high in soluble dietary fiber, has a balanced amino acid content, and contains high levels of Vitamins A and C <sup>5</sup>. Teff is mostly grown by small scale farmers for personal consumption and to be sold at market, and it accounts for ~10% of the total calories consumed in Ethiopia <sup>6</sup>. Teff is most commonly milled to a flour and used in the fermented flatbread injera which has been reported to maintain high concentrations of Fe, Ca, and Zn <sup>7</sup>.

Globally, teff is a nutritious and gluten free alternative to wheat flour and has seen sustained growth for use in other food products including breads, pastas, porridge, and malt for brewing, as well as fodder for animals.

Previous research on teff grains has highlighted its rich nutritional profile. However nutritional studies in teff show considerable variability in seed mineral content, typically reporting on only a few genotypes with limited genetic variability, paired with data collected from a single environment or year. These findings also differ in methodology, and often report concerns of soil contamination. Micro and macronutrient profiles are known to be impacted by the environment with high genotype by environmental variance. In Ethiopia, nutritional quality of teff varied significantly based on planting location<sup>8</sup>. Compared to maize and barley, high and variable iron contents were reported in teff seed and injera<sup>9</sup>. Additionally, it has been reported that teff seed nutrient content varies by seed color with numerous reports of brown or red seeded teff being higher in Fe, Zn, Ca<sup>10</sup>. White seed is currently preferred and sold at a higher price, therefore a comprehensive quantification of white and brown seed nutritional differences could impact teff market production and economics.

Descriptive research of teff nutrient transport, storage, and accumulation is limited. However, gene families involved in metal transport were recently evaluated across 24 teff varieties<sup>11</sup>. With the development of molecular markers and potential regions of interest associated with teff seed nutrient concentrations we can begin to piece together the underlying genetic mechanisms of teff nutrient metabolism and how this resilient grain has maintained its nutritional stability over time. With the tools of marker-assisted breeding we can work toward biofortification of teff and also apply our findings to the improvement of other cereals. Ultimately harnessing the variability of teff germplasm we can rapidly improve teff varietal development.

In this study we report the first multi-year field trial phenotyping macro and micro nutrient seed concentrations across a diverse panel of 265 teff accessions from nine subpopulations as well as the wild progenitor *Eragrostis pilosa*. The Teff Association Panel (TAP) was previously sequenced and contains a large proportion of Ethiopian teff genetic diversity. Paired with phenotypic data for 12 seed nutrients, we conducted genome wide association studies and identified 19 loci from Phosphorus, Potassium, Zinc, Iron, and Aluminum. The seed nutrient variability and regions associated with these loci can be used by

breeders, scientists, and growers to select varieties, and enhance nutritional improvement of teff.

## **Methods**

### ***Plant Materials and field conditions***

Seed nutritional content was surveyed using the Teff Association Panel (TAP), which includes teff landraces and cultivars across nine subpopulations as well as teff's wild progenitor, *Eragrostis pilosa*. The panel was grown in triplicate using a randomized block design in the summer of 2021 and 2022 with single row, 4.5 ft plots, at the Michigan State University Horticulture Teaching and Research Center (HTRC) in Holt, MI (42°67'43.4"N, 84°48'43.5"W). Approximately 100 teff seeds of each accession were planted by hand in each plot on June 1st in 2021 and June 3rd, 2022. To enhance yield and uniformity, a fertilizer treatment of 19-19-19 at ~100 lbs/A was applied prior to planting. Additionally, herbicide was applied to the entire field for the control of broadleaves (Broclean), and between rows for grasses (Roundup PowerMAX). Soil health is maintained by the HTRC, and soil cores were sampled at the end of each season across the field and presented optimum phosphorus, potassium, and magnesium levels with slightly high pH of 6.8 and 7.6 on average in 2021 and 2022 (**Table 3.1**). Plots were harvested by hand November 2nd-10th, 2021 and October 20-28th, 2022.

### ***Trait phenotyping***

Panicles were harvested at maturity and dried in paper bags. Seed was hand threshed from the panicle, sieved, and aspirated with the Seedburo 757 South Dakota Seed Blower. Approximately one gram of cleaned seed from each plot was manually inspected for debris and sent to A&L Great Lakes Laboratory for Inductively Coupled Plasma Optical Emission spectroscopy (ICP-OES) to test micro and macro nutrient concentrations (P1DRY) within the samples. Phosphorus, Potassium, Calcium, Magnesium, Sodium, Sulphur, Zinc, Manganese, Iron, Copper, Boron, and Aluminum seed sample concentrations were reported. The poor conditions in 2022 were reflected in our sample collection. In 2021 we sampled 261 accessions, and 86% of the samples had three replicates across three blocks represented. In 2022, however, we only sampled 194 accessions and only 50% of the samples had three blocks represented in our final dataset. Seed was classified as brown, white, or mixed in color then 185 accessions with consistent white and brown seed color were selected for further analysis.

### ***Statistical analysis of field data***

Statistical analyses were performed in R software version 4.3.3 unless otherwise noted.

An ANOVA was carried out to model the genetic, block, year, and subpopulation effects on each trait. The Best Linear Unbiased Estimators (BLUES) for each trait were estimated using lme4 where block and year were fit as random effects and accession was fit as a fixed effect for each trait, since plots were planted in a randomized complete block design. BLUES of each accession for each trait were used as phenotype data for GWA. Broad-sense heritability ( $h^2$ ), or in this case repeatability, was estimated as a secondary model in which accession, block, and year were fit as random effects to estimate variance components via restricted maximum likelihood (REML) using lme4.  $H^2$  values of the traits were estimated according to  $h^2 = \frac{\sigma^2}{\sigma^2 + \sigma e^2}$ , where  $\sigma^2$  is the genotypic variance and  $\sigma e^2$  is the residual variance. Pearson's correlation of BLUES of each trait was calculated using the corr() function in the Python package pandas (v2.2) and employed a Mann-Whitney U Test to examine the effect of seed color on nutrition using scipy.stats (statsmodels v0.14.2) <sup>12 13</sup>.

### ***Genome wide association***

Whole genome resequencing data was previously generated for the TAP, and we used a subset of LD pruned single nucleotide polymorphisms as genetic markers for genome wide association (GWA). In total, 746,242 LD pruned, imputed SNP based markers were used. Kinship relatedness matrix and principal component analysis on genetic structure of the panel were conducted in TASSEL <sup>14</sup>. The first three principal components and kinship matrix were applied as covariates to account for population structure and relatedness of the panel. GWA for Phosphorus, Potassium, Calcium, Magnesium, Sulphur, Zinc, Manganese, Iron, Copper, Boron, and Aluminum were conducted using GAPIT version 3.4 in R with trait BLUES <sup>15</sup>. GWA was performed using a range of models including general linear model (GLM), mixed linear model (MLM), compressed MLM (CMLM), multiple loci mixed model (MLMM), Fixed and random model Circulating Probability Unification (FarmCPU), and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway BLINK. Model performance was evaluated using quantile-quantile plots and based on model fit, the BLINK model was chosen for further analysis. P-values for each SNP were visualized as Manhattan plots with QQMan <sup>16</sup>. Candidate genes for each trait were identified based on their proximity of 100 kb up- and downstream of significant markers identified in the GWA. Candidate genes were annotated based on KEGG and GO terms to describe generalized metabolic and cellular processes responses. KEGG annotations were generated for each SNP using BLASTKoala (<https://www.kegg.jp/blastkoala/>) and were

used to create metabolic pathway maps with KEGGmapper (<https://www.genome.jp/kegg/mapper/color.html>).

## Results

### *Variability of seed nutrient concentrations*

We estimated concentrations of micro and macronutrients across the TAP from field grown samples collected in 2021 and 2022. To minimize contamination, we utilized a standardized cleaning and analysis protocol, using plastic and stainless steel whenever possible. Samples were harvested, dried, threshed, aspirated, and then subjected to Inductively Coupled Plasma Optical Emission spectroscopy (ICP-OES). Data was collected for 12 nutrient concentrations. Phosphorous, Potassium, Calcium, Magnesium, Sodium, and Sulphur (P, K, Ca, Mg, Na, S) were reported as percent dm, and Zinc, Manganese, Iron, Copper, Boron, and Aluminum (Zn, Mn, Fe, Cu, B, Al) were reported as ppm dm. The 2021 field season had much higher cumulative rainfall, resulting in higher, more consistent yields compared to 2022. This is reflected in the data as an increase of sample size in 2021, as well as higher nutritive values in 2021 for P, Mg, S, Zn, Mn, Cu, and B.

We observed a normal distribution and relatively low variation in concentrations of P, K, Ca, Mg, S, Zn, Cu, and B with coefficients of variation (CV) ranging from 7.44-18.8 and no variation in Na concentrations (**Figure 3.1, Table 3.2**). Manganese has high variability across the panel (CV=58.3) and only 60 accessions have a CV less than 30. Iron is also highly variable (CV= 121) and only 141 accessions have a CV less than 30. When separated by year however, 234 accessions have a CV<30 in 2021, compared to only 64 in 2022. Aluminum has the highest variability among samples (CV=292), and only 13 accessions have a CV of less than 30. This pattern is consistent across years with 46 and 13 accessions maintaining CV<30 in 2021 and 2022, respectively. The substantial variability observed in Manganese, Iron, and Aluminum suggests greater fluctuations compared to other nutrients. Despite the known influence of environmental factors on seed nutrient content, we report particularly drastic annual variations in the consistency of teff iron content.

Wild species often have higher nutritional contents than their domesticated relatives, and this is frequently attributed to differences in seed size which is correlated with mineral content due to endosperm composition <sup>17–20</sup>. Teff seeds are considerably smaller than other cereal crops, and we investigated if teff is less nutritious than its small-seeded wild relative, *E. pilosa*. We

harvested sufficient seed for four *E. pilosa* accessions. The results indicated that *E. pilosa* possessed the highest nutritional content among all analyzed traits, except for Zinc (**Figure 3.2**).

To evaluate the significance of genotypic and environmental effects, we performed ANOVA for each trait. The environmental effects of year and block were significant across each trait, except for Potassium and Boron which showed no significant difference between years (**Table 3.3**). Additionally, subpopulation was not significant for Mn, Fe, B, or Al. Each trait we measured across the TAP had low broad-sense heritability, or repeatability, ranging from 0.01 for Boron to 0.54 for Phosphorus (**Table 3.2**). The low heritabilities of these traits emphasize the importance of environmental impact on teff seed nutrient content.

To adjust for environmental variability and minimize bias across samples, we calculated Best Linear Unbiased Estimators (BLUES) and used these values to assess the correlations among traits (**Figure 3.3**). The majority of the micronutrients are significantly positively correlated ( $p$  value  $<0.05$ ). Phosphorus is strongly correlated with Magnesium ( $r=0.66$ ,  $pval=1.59e-34$ ) and Sulfur ( $r=0.57$ ,  $pval=5.75e-24$ ). Sulphur has a strong correlation with Zinc ( $r=0.56$ ,  $pval=1.173e-23$ ) and Copper ( $r=0.63$ ,  $pval=8.34e-31$ ), and Zinc is highly correlated with Copper ( $r=0.58$ ,  $pval=1.94e-25$ ). The strongest correlation is that of Iron and Aluminum ( $r=0.92$ ,  $pval=7.37e-112$ ).

#### ***Testing for associations between seed color and mineral nutrition***

To test the effect of seed color on nutrition, a subset of 185 samples with consistently colored seed was evaluated using Mann-Whitney U Test (**Figure 3.4**). Except for Boron, all nutrients had significant differences between brown and white seed. Six nutrients, Zn, Cu, P, K, Mg, and S, had higher means in white seeded individuals. Conversely, brown seeded accessions have a higher average Mn, Fe, Al, and Ca nutrient contents.

#### ***Detecting genetic loci related to seed macro and micronutrients***

We performed genome-wide association (GWA) and employed multiple models to detect genetic loci significantly associated with seed mineral nutrient concentrations in 259 accessions of the TAP. Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) models displayed the best fit, and robust QQ-plots for P, K, Zn, Fe, and Al. Unlike FarmCPU, BLINK does not evaluate loci through evenly distributed bins, but instead uses linkage disequilibrium to identify potential QTL <sup>21</sup>. BLUES calculated from the two field seasons served as phenotypic data and covariates including kinship and the first three principal

components from the TAP genetic diversity analysis were included.

Marker-trait associations were identified for five out of the eleven nutrients sampled: P, K, Zn, Fe, and Al. We identified a total of 19 loci significantly associated with these traits (**Table 3.5, Figure 3.5-3.6**). Three loci were detected on Chromosomes 10A, 1A 1B, 5A, and 9A, followed by two loci on Chromosome 2B and one on Chromosome 3B. It was previously reported that subgenome A has a higher nucleotide diversity and a greater range of gene expression than subgenome B, so a higher number of significant loci in subgenome A was expected <sup>22</sup>.

#### ***Significant loci associated with seed nutrient concentrations***

For Potassium, four loci were discovered on Chromosomes 3B, 4A, 9A, and 10A. The SNP on Chromosome 10A (4,042,845 bp) explains 33.8% of phenotypic variance (PVE) and the additional three loci explain 2.73-14.4%. The SNP on Chromosome 10A lies within the 5'UTR region of Et\_10A\_001497 which shares significant sequence identity with disease resistance genes. Plants with high potassium are reportedly less susceptible to disease <sup>23</sup>. Five loci were significantly associated with Phosphorus on Chromosomes 1A, 1B, 2B, and 5A and explained 3.87-26.1% of phenotypic variance. Chromosome 1B (9,910,062 bp) had the highest PVE, upon review, genes surrounding this marker are key players in gene regulation including an array of transcription factors and binding domains potentially involved in nutrient metabolism, transport, and storage. Two loci were identified for Zinc, Chromosome 1A (39,734,229) (PVE=8.32 %) and Chromosome 10A (14,502,337) (PVE=52.6%).

Iron content is affected by six loci across Chromosomes 1B, 5A, 9A, and 10A, with PVE ranging from 4.92% to 41.1%. The locus on Chromosome 5A at 21,796,480 bp, which is located just 43Kb from another significant hit, displayed the highest PVE, indicating a region rich in genetic determinants of iron content. Protein domains such as F-box, NB-Arc, and Pentatricopeptide repeat, detected in this region, are prime candidates for further investigation.

Three loci associated with Aluminum content were detected on Chromosomes 1A, 1B, and 2B, explaining 12.3% to 44.17% of PVE. The highest PVE was observed on Chromosome 2B at 9,954,696 bp. The gene Et\_2B\_022156, located 34Kb away, encodes a b-Zip transcription factor known for their interaction with seed-specific proteins affecting endosperm, oil, and nutrient content <sup>24-26</sup>. Interestingly, one locus on Chromosome 1B at 8,458,558 bp influences both Aluminum and Iron content with PVEs of 22.4% and 6.66%, respectively. This correlation

aligns with the observed high interdependence between these two nutrients.

## **Discussion**

Here we report the first large scale dataset of teff seed nutrient content. Teff is consistently reported as being a nutritious grain, but this has not been quantified for diverse accessions or across multiple environments. Although this study is limited to teff performance in Michigan, our estimations of seed mineral nutrition can be pivotal for future teff selection, improvement, and nutritional analysis.

Previous reports of teff seed nutrient content are inconsistent and often only characterize a small panel of varieties. One study sampled three teff varieties of differing color from teff producing regions, Bahir Dar, Debre Markos, and Bure and found extremely high levels of Iron and Aluminum >1000 mg/kg (or ppm dm) using microwave plasma-atomic emission spectrometry (MPAES) <sup>27</sup>. These levels were only reported by a single replicate of PI494209 and PI494188 in our study in 2022. When we compare our results for Zn and Cu, they report much higher concentrations as well with mean values of 69-102 and 13-15 respectively; however, they report similar levels of Mn. It's notable that teff is commonly milled into flour for analysis, primarily because it is predominantly consumed as the fermented flatbread injera. The largest panel previously surveyed consists of 24 teff varieties, and flour nutrition was evaluated via inductively coupled plasma mass spectrometry (ICP-MS). Zn flour concentrations ranged from 14.8-29.2 mg/kg and Fe concentrations were 22.6-684.25 mg/kg <sup>11</sup>. These findings are in line with our averages; however, direct comparisons remain challenging due to variations in sample preparation and analytical methodologies.

The controversy of teff seed nutrient contamination emerged due to the combination of small seeded grain and traditional threshing practices. Teff is most commonly threshed on the ground, introducing soil contaminants that easily coat the small seed <sup>28-30</sup>. To avoid potential contamination, we employed a standardized protocol for threshing and sample processing consisting of plastic and stainless steel when available. Although we have high variability in teff Mn, Fe, and Al concentrations, we believe these values accurately reflect the panel's diversity.

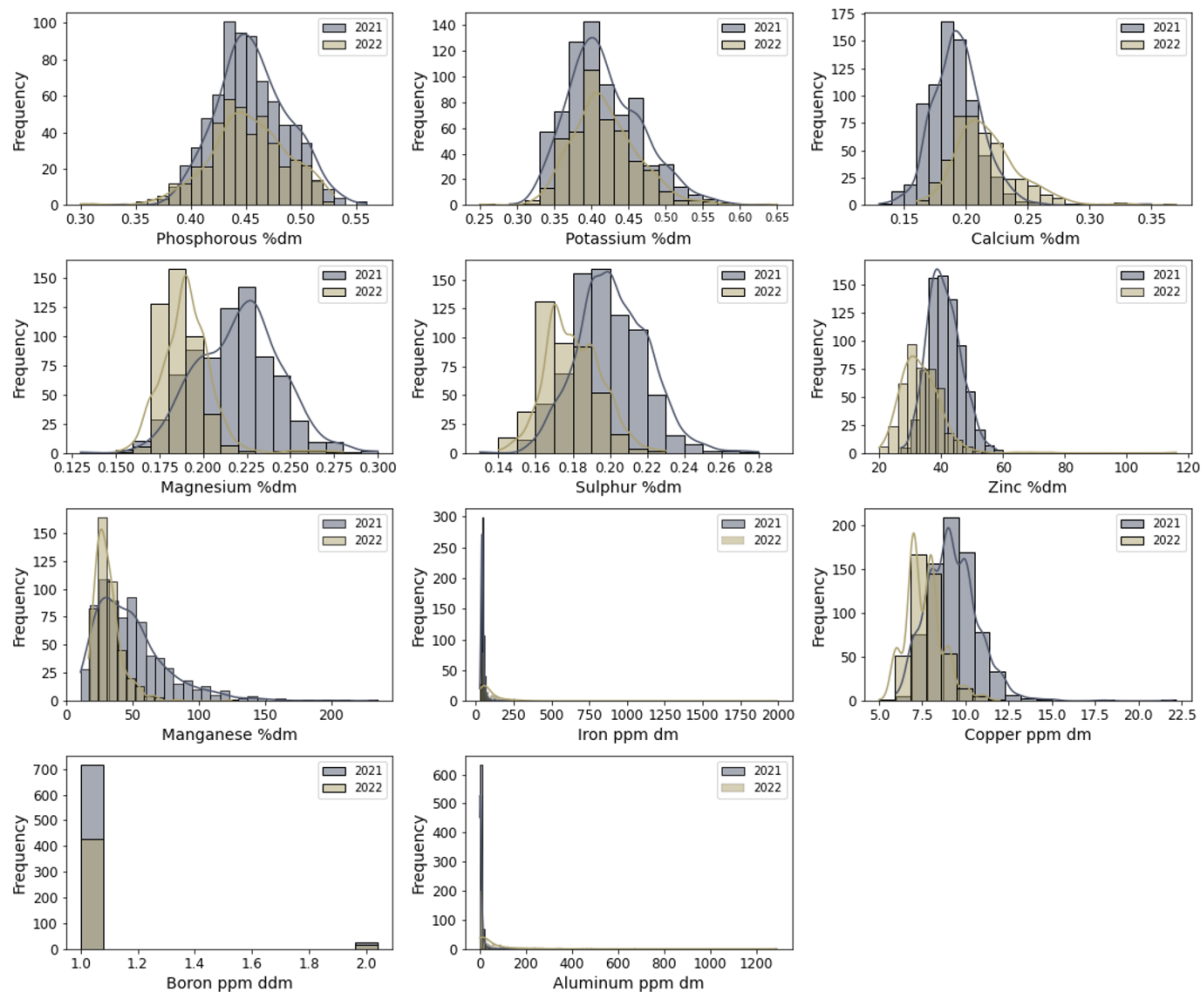
To begin isolating key players in teff nutrient transport and seed storage regulation, this work was supported by GWA. We identified 19 loci for further review. A detailed examination of the genes surrounding these loci revealed numerous binding domains and transcription factors potentially impacting the expression and regulation of mineral content, including F-box, zinc



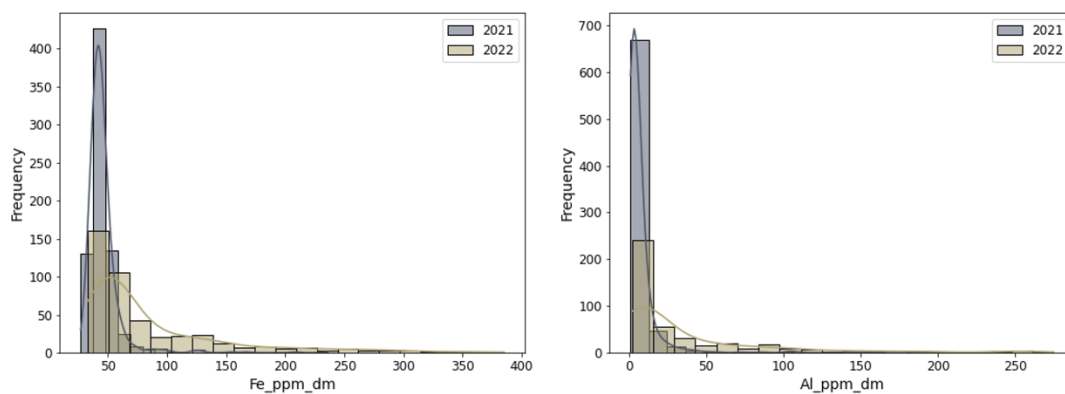
finger, and bZIP. Additionally, Gene Ontology (GO) terms related to trehalose, lysine, de novo UMP, spermidine, spermine, de novo pyrimidine nucleobase, and amino acid biosynthetic processes, chitin catabolic process, tRNA wobble uridine modification, and isocitrate metabolic processes were enriched near significant loci, and can point to mechanisms of seed mineral accumulation (**Table 3.6**). This study not only unveils the first large-scale dataset of teff seed nutrient content but also highlights the influence of genetic and environmental factors on nutrient variability, paving the way for targeted genetic improvements and enhanced understanding of teff's nutritional benefits.

## Figures and Tables

A



B

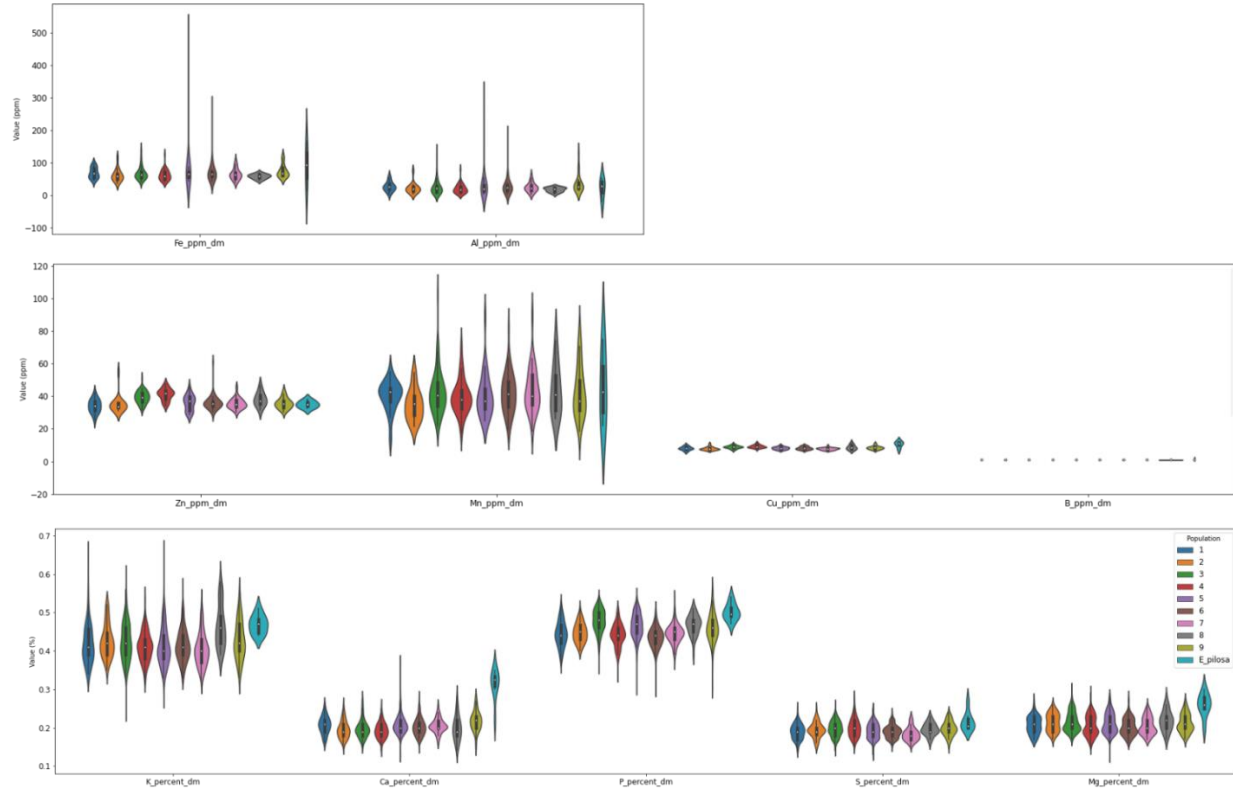


**Figure 3.1: Phenotypic distribution of seed nutrient content within TAP in 2021 and 2022.**

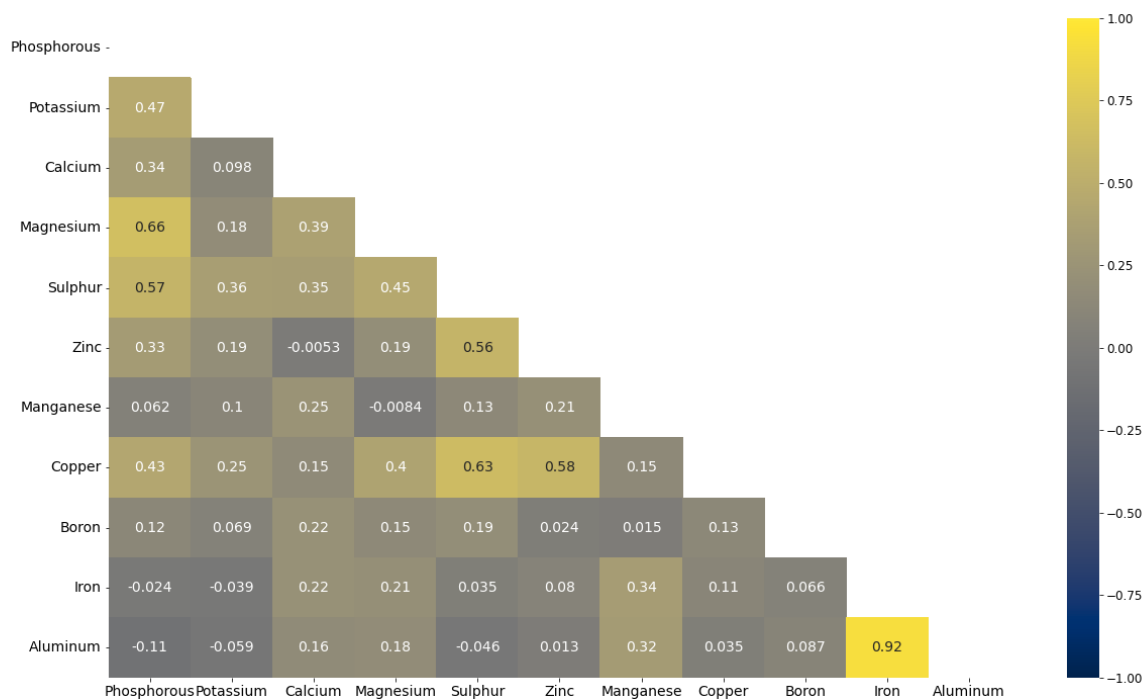
A) Distribution of phosphorous, potassium, calcium, magnesium, sulphur, zinc, manganese, iron,

**Figure 3.1 (cont'd)**

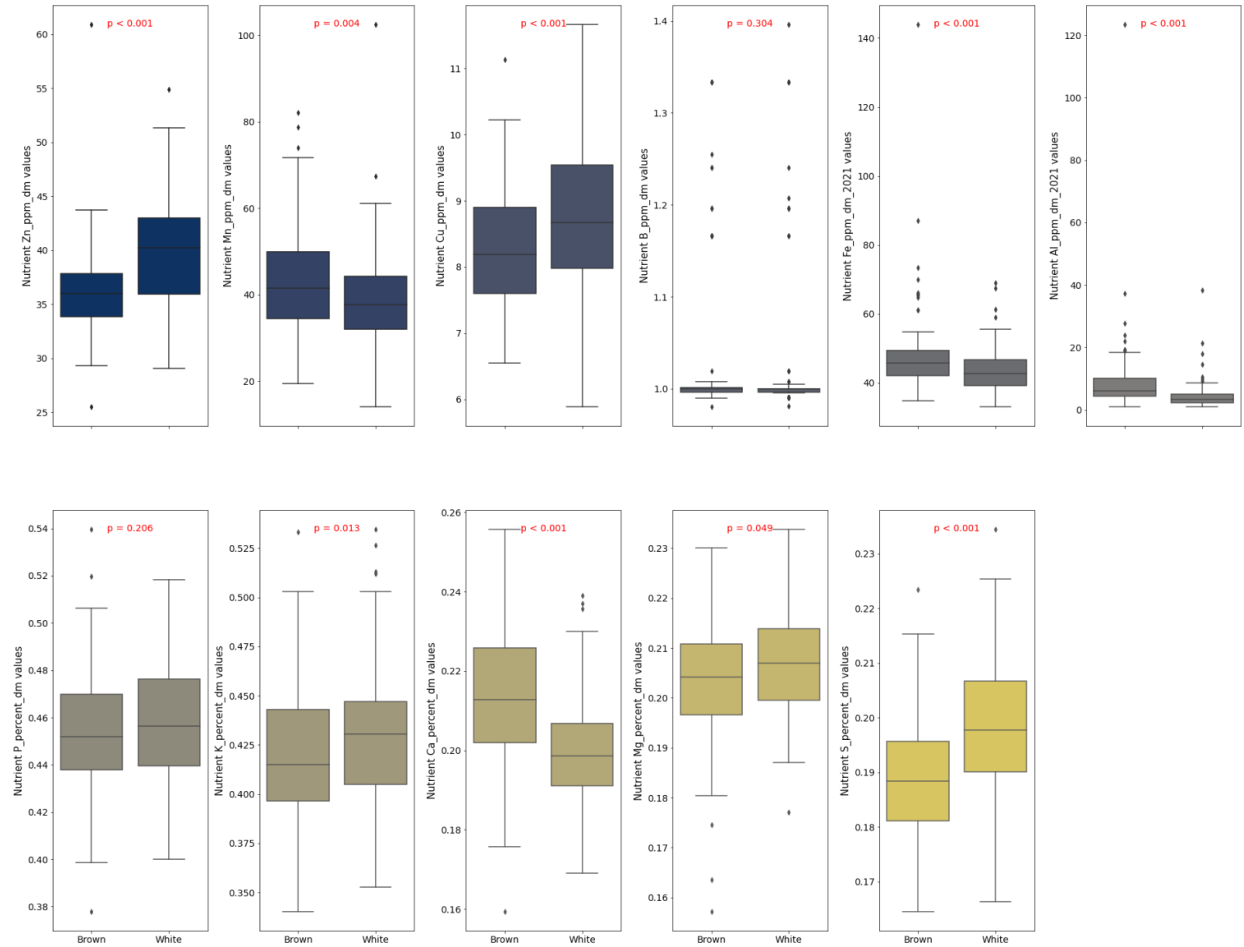
copper, boron, and aluminum. B) Subset of values for iron and aluminum to more clearly visualize the distribution trend.



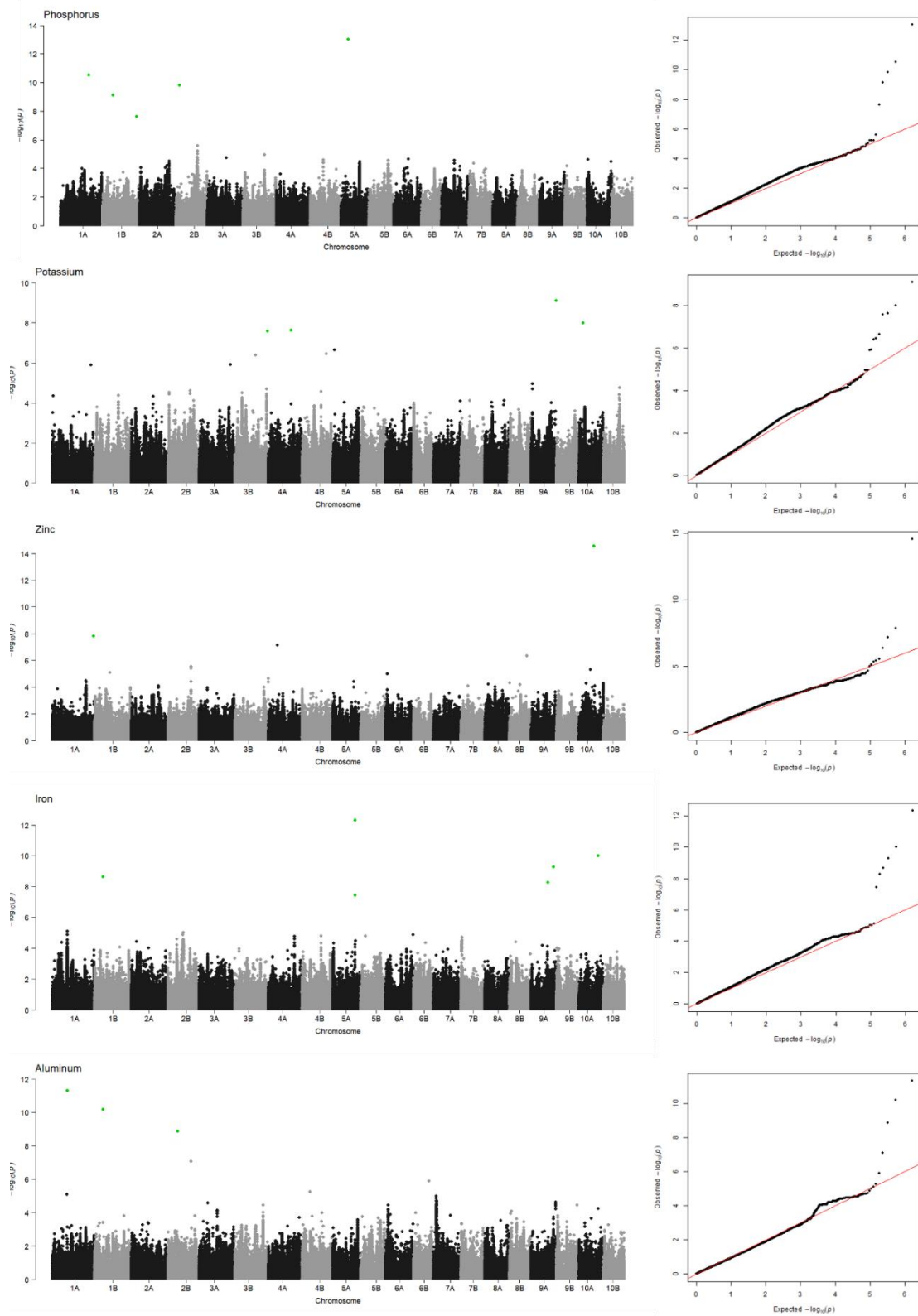
**Figure 3.2: Distribution of micronutrient concentrations across subpopulations.**



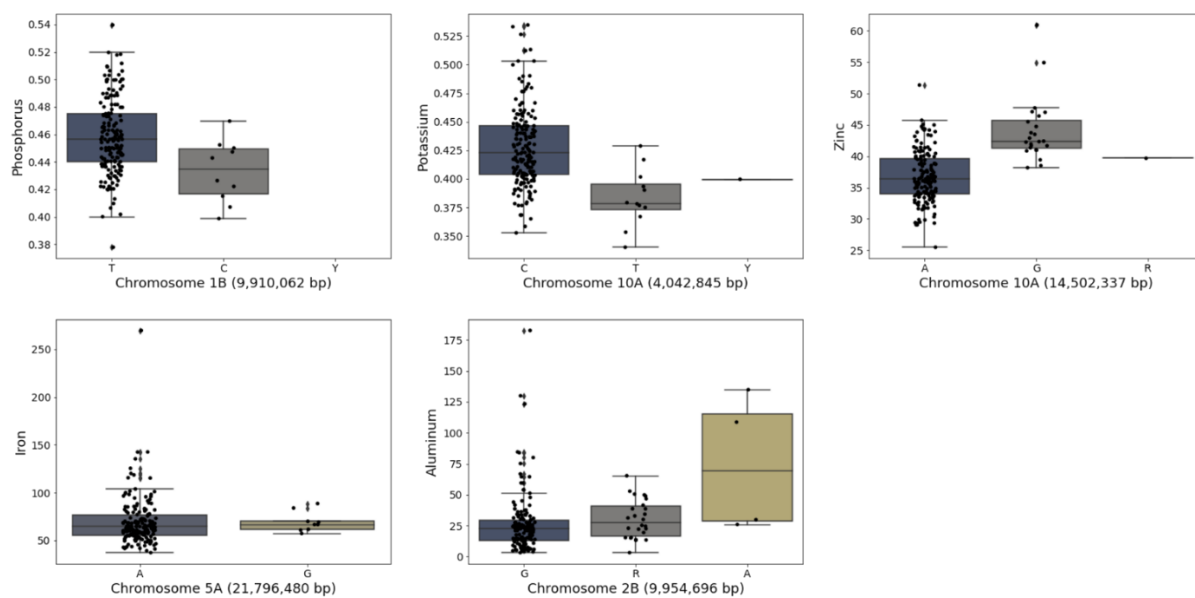
**Figure 3.3: Correlation between seed nutrient content BLUES.**



**Figure 3.4: Distribution of micronutrient concentrations across brown and white seeded individuals.**



**Figure 3.5: BLINK GWA Manhattan and QQ plots for days to phosphorus, potassium, zinc, iron, and aluminum, significant loci highlighted in green.**



**Figure 3.6: Allele frequency of most significant loci discussed for phosphorus, potassium, zinc, iron, and aluminum**

**Table 3.1:** Soil core samples from MSU diagnostics laboratory, sample IDs are labeled as 1-3 corresponding to blocks within the field.

<b>Year</b>	<b>Sample I.D.</b>	<b>Ph</b>	<b>P</b>	<b>K</b>	<b>Ca</b>	<b>Mg</b>	<b>Zn</b>	<b>Mn</b>	<b>Cu</b>	<b>Fe</b>	<b>B</b>	<b>S</b>
<b>2021</b>	1a	6.4	36	143	1188	160	1.9	15.0	3.7	53	0.2	10.0
<b>2021</b>	1b	7.2	21	113	1776	224	1.7	23.8	3.0	45	0.4	8.0
<b>2021</b>	1c	6.6	21	124	1370	190	1.7	20.3	2.8	47	0.5	11.0
<b>2021</b>	2a	6.2	55	153	1543	197	3.2	21.6	4.5	75	0.4	13.0
<b>2021</b>	2b	7.0	44	183	1809	264	4.9	28.9	4.8	48	0.6	10.0
<b>2021</b>	2c	6.5	43	173	1521	196	4.4	22.7	3.6	52	0.5	9.0
<b>2021</b>	3a	7.2	53	192	2025	286	4.5	32.0	6.3	52	0.8	12.0
<b>2021</b>	3b	6.8	52	212	1855	258	3.6	24.7	6.4	55	0.6	12.0
<b>2021</b>	3c	7.0	46	184	1947	252	2.5	25.4	5.9	48	0.7	12.0
<b>2022</b>	1a	7.8	55	146	1301	210	1.2	24.8	2.9	100	0.4	3
<b>2022</b>	1b	7.5	63	186	1223	216	1.4	25.2	3.4	100	0.3	3
<b>2022</b>	1c	7.4	55	191	1158	204	1.1	21.1	2.7	91	0.3	2
<b>2022</b>	2a	7.5	52	158	1215	198	1.1	24.6	3.2	93	0.3	1
<b>2022</b>	2b	7.4	61	187	112	198	1.3	24.4	2.8	92	0.3	2
<b>2022</b>	2c	7.5	61	199	1183	198	1.4	26.3	3.1	93	0.3	3
<b>2022</b>	3a	7.1	54	166	1105	169	1	17.6	2.8	58	0.3	3
<b>2022</b>	3b	7.1	69	198	1029	164	1.3	21.6	3.7	58	0.3	2
<b>2022</b>	3c	7.6	59	298	1149	216	1.4	27.2	3.4	69	0.3	2



**Table 3.2:** Phenotypic diversity metrics.

Trait	Year	Unit	Median	Range	Mean ± Sd	CV	h <sup>2</sup>	σ <sup>2</sup> <sub>g</sub>	σ <sup>2</sup> <sub>r</sub>
Phosphorous	2021- 2022	percent dm	0.45	0.3- 0.56	0.46 ± 0.034	7.44	0.54	0.000623	0.00053
Potassium	2021- 2022	percent dm	0.41	0.25- 0.65	0.42 ± 0.048	11.5	0.37	0.000900	0.00134
Calcium	2021- 2022	percent dm	0.20	0.13- 0.37	0.20 ± 0.026	12.8	0.40	0.000364	0.00022
Magnesium	2021- 2022	percent dm	0.21	0.13- 0.3	0.21 ± 0.026	12.2	0.14	0.000139	0.00020
Sulfur	2021- 2022	percent dm	0.19	0.13- 0.28	0.19 ± 0.020	10.5	0.24	0.000126	0.00018
Zinc	2021- 2022	ppm dm	38.0	20-116	38.2 ± 7.17	18.8	0.25	16.5	21.5
Manganese	2021- 2022	ppm dm	35.0	11-233	42.7 ± 24.9	58.3	0.06	40.1	458
Iron	2021- 2022	ppm dm	46.0	27- 1993	63.2 ± 76.6	121	0.04	237	5058
Copper	2021- 2022	ppm dm	8.00	5-22	8.64 ± 1.59	18.4	0.30	0.96	0.97
Boron	2021- 2022	ppm dm	1.00	1-2	1.04 ± 0.185	17.9	0.01	0.00034	0.03318

**Table 3.2 (cont'd)**

Aluminum	2021- 2022	ppm dm	5.00	1-1288	21.4 ± 62.6	292	0.04	177	3315
Iron	2021	ppm dm	43.00	27237- 236	45.65 ± 15.6	34.1	0.45	53.59307	193.4621
Aluminum	2021	ppm dm	4.00	1-	6.75 ±14.0	206	0.10	6.892127	188.7564

**Table 3.3:** ANOVA results from linear models across agronomic traits, significance codes: 0 \*\*\*0.001, \*\*0.01, \*0.05.

Trait	Year Pvalue	Block Pvalue	Subpopulation Pvalue	Accession Pvalue
Phosphorous	0.01994*	0.00612**	<2.2e-16 ***	<2.2e-16 ***
Potassium	0.9638	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***
Calcium	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***
Magnesium	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***
Sulfur	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***	<2.2e-16 ***
Zinc	<2.2e-16 ***	9.095e-09***	<2.2e-16 ***	<2.2e-16 ***
Manganese	<2.2e-16 ***	<2.2e-16 ***	0.1043	7.554e-05***
Iron	<2.2e-16 ***	0.0002178 ***	0.1512921	0.1564052
Copper	<2.2e-16 ***	2.282e-05***	<2.2e-16 ***	<2.2e-16 ***
Boron	0.8268	5.83e-07***	0.2024	0.1032
Aluminum	<2.2e-16 ***	2.265e-06 ***	0.10511	0.07921

**Table 3.4:** Significant GWA loci.

SNP	Chrom	Position	P.value	MAF	Effect	PVE (%)	Trait
Chromosome_1A_14521895	1	14521895	4.67E-12	0.069767	16.91948	13.29103	Al
Chromosome_1A_39734229	1	39734229	1.44E-08	0.267442	-1.42711	8.317669	Zn
Chromosome_1A_27452084	1	27452084	2.93E-11	0.168605	-0.01135	10.81432	P
Chromosome_1B_8458558	2	8458558	6.38E-11	0.04845	-18.939	22.43931	Al
Chromosome_1B_8458558	2	8458558	2.18E-09	0.04845	-20.8076	6.666651	Fe
Chromosome_1B_9910062	2	9910062	7.20E-10	0.052326	0.014448	26.13589	P
Chromosome_1B_33162248	2	33162248	2.27E-08	0.436047	0.005639	3.868816	P
Chromosome_2B_9954696	4	9954696	1.34E-09	0.096899	-17.1013	44.17486	Al
Chromosome_2B_3584563	4	3584563	1.54E-10	0.065891	0.013723	23.87844	P
Chromosome_3B_31671879	6	31671879	2.54E-08	0.358527	0.008875	2.726125	K
Chromosome_4A_22229833	7	22229833	2.33E-08	0.129845	0.012125	11.08945	K
Chromosome_5A_21796480	9	21796480	4.70E-13	0.054264	23.58208	41.14551	Fe
Chromosome_5A_21840213	9	21840213	3.55E-08	0.063953	-36.7763	16.24104	Fe
Chromosome_5A_7143120	9	7143120	9.12E-14	0.292636	-0.00942	15.59589	P
Chromosome_9A_16154088	17	16154088	5.26E-09	0.052326	25.80787	20.76379	Fe
Chromosome_9A_21888521	17	21888521	5.27E	0.04845	-	5.011789	Fe

**Table 3.4 (cont'd)**

			-10		21.0387		
Chromosome_9A_24138084	17	24138084	7.84E	0.255814	0.01010	14.37878	K
			-10		9		
Chromosome_10A_1450233	19	14502337	2.75E	0.110465	4.14957	52.55093	Zn
7			-15		3		
Chromosome_10A_1861587	19	18615876	9.90E	0.056202	-	4.920378	Fe
6			-11		21.2498		
Chromosome_10A_4042845	19	4042845	9.77E	0.062016	-	33.75626	K
			-09		0.01582		

**Table 3.5:** Significant GO term associations with pval<0.05 for genes +- 100Kb from significant loci.

GO.ID	Term	Annotated	Significant	Expected	classic Fisher	Genes
GO:0005992	trehalose biosynthetic process	41	4	0.28	0.00016	Et_10A_000880, Et_1A_006770, Et_9A_062572, Et_9A_062575
GO:0009089	lysine biosynthetic process via diaminop...	15	2	0.1	0.0045	Et_1A_005829, Et_1A_005839
GO:0006032	chitin catabolic process	24	2	0.16	0.01138	Et_9A_063222, Et_9A_063223
GO:0044205	'de novo' UMP biosynthetic process	2	1	0.01	0.01355	Et_3B_030244
GO:0002098	tRNA wobble uridine modification	2	1	0.01	0.01355	Et_10A_001502
GO:0009082	branched-chain amino acid biosynthetic p...	28	2	0.19	0.01531	Et_1B_013449, Et_9A_062315
GO:0008295	spermidine biosynthetic process	4	1	0.03	0.02691	Et_1B_013442
GO:0006424	glutamyl-tRNA aminoacylation	6	1	0.04	0.0401	Et_1A_007963
GO:0006597	spermine biosynthetic process	6	1	0.04	0.0401	Et_1B_013442

**Table 3.5 (cont'd)**

GO:0006102	isocitrate metabolic process	6	1	0.04	0.0401	Et_9A_062323
GO:0006207	'de novo' pyrimidine nucleobase biosynth...	7	1	0.05	0.04663	Et_3B_030244

## REFERENCES

1. Mantri, N., Patade, V., Penna, S., Ford, R. & Pang, E. Abiotic Stress Responses in Plants: Present and Future. *Abiotic Stress Responses in Plants* 1–19 Preprint at [https://doi.org/10.1007/978-1-4614-0634-1\\_1](https://doi.org/10.1007/978-1-4614-0634-1_1) (2012).
2. Rippey, B. R. The U.S. drought of 2012. *Weather and Climate Extremes* vol. 10 57–64 Preprint at <https://doi.org/10.1016/j.wace.2015.10.004> (2015).
3. International Food Policy Research Institute (IFPRI). *Global Nutrition Report. 2022 Global Nutrition Report: Stronger commitments for greater action*. Bristol, UK: Development Initiatives, 2022.
4. Baye & Kaleab. *Teff: Nutrient Composition and Health Benefits*. (Intl Food Policy Res Inst).
5. Yilmaz, H. O. & Arslan, M. Teff: Nutritional compounds and effects on human health. *Acta Sci. Med. Sci* **2**, 15–18 (2018).
6. Wang Y, Çakır M. Welfare impacts of increasing teff prices on Ethiopian consumers. *Agricultural Economics*. **52**, 195–21 (2021).
7. Comprehensive study on the effect of fermentation time, baking temperature and baking time on the physicochemical and nutritional properties of injera teff (*Eragrostis teff*). *Food and Humanity* **2**, 100256 (2024).
8. Gashu, D. *et al.* The nutritional quality of cereals varies geospatially in Ethiopia and Malawi. *Nature* **594**, 71–76 (2021).
9. Abebe, Y. *et al.* Phytate, zinc, iron and calcium content of selected raw and prepared foods consumed in rural Sidama, Southern Ethiopia, and implications for bioavailability. *J. Food Compost. Anal.* **20**, 161–168 (2007).
10. Lamesgen, Y *et al.* Comparative analysis of proximate and mineral composition of released Tef (*Eragrostis tef* (Zucc.) Trotter) varieties in Ethiopia. *Academic Research Journal of Agricultural Science and Research*. 372-379 (2019)
11. Ereful, N. C. *et al.* Nutritional and genetic variation in a core set of Ethiopian Tef (*Eragrostis tef*) varieties. *BMC Plant Biol.* **22**, 1–14 (2022).
12. statsmodels 0.15.0 (+263). <https://www.statsmodels.org/devel/>.
13. Website. The pandas development team. (2024). pandas-dev/pandas: Pandas (v2.2.2). Zenodo. <https://doi.org/10.5281/zenodo.10957263>.
14. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).

15. GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics Proteomics Bioinformatics* **19**, 629–640 (2021).
16. D. Turner, S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
17. Cakmak, I., Ozkan, H., Braun, H. J., Welch, R. M. & Romheld, V. Zinc and Iron Concentrations in Seeds of Wild, Primitive, and Modern Wheats. *Food Nutr. Bull.* (2000) doi:10.1177/156482650002100411.
18. Bálint, A. F., Kovács, G., Erdei, L. & Sutka, J. Comparison of the Cu, Zn, Fe, Ca and Mg contents of the grains of wild, ancient and cultivated wheat species. *Cereal Res. Commun.* **29**, 375–382 (2001).
19. NUTRITIONAL QUALITY OF SOME CULTIVATED AND WILD SPECIES OF AMARANTHUS L. *INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH / IJPSR* <https://ijpsr.com/bft-article/nutritional-quality-of-some-cultivated-and-wild-species-of-amaranthus-l/> (2011).
20. Swamy, B. P. M., Marathi, B., Ribeiro-Barros, A. I. F. & Ricachenevsky, F. K. Editorial: Development of Healthy and Nutritious Cereals: Recent Insights on Molecular Advances in Breeding. *Front. Genet.* **12**, 635006 (2021).
21. Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **8**, giy154 (2018).
22. VanBuren, R. *et al.* Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. *Nat. Commun.* **11**, 884 (2020).
23. Tripathi, R. *et al.* Plant mineral nutrition and disease resistance: A significant linkage for sustainable crop protection. *Front. Plant Sci.* **13**, (2022).
24. Li, Y. *et al.* Transcriptome and metabolome reveal distinct carbon allocation patterns during internode sugar accumulation in different sorghum genotypes. *Plant Biotechnol. J.* **17**, 472 (2019).
25. Weltmeier, F. *et al.* Expression patterns within the Arabidopsis C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol. Biol.* **69**, 107–119 (2008).
26. Cifuentes-Esquivel, N. *et al.* bZIP17 regulates the expression of genes related to seed storage and germination, reducing seed susceptibility to osmotic stress. *J. Cell. Biochem.* **119**, 6857–6868 (2018).
27. Gebregewergis, A., Chandravanshi, B. S. & Redi-Abshiro, M. Levels of selected metals in teff grain samples collected from three different areas of Ethiopia by microwave plasma-atomic emission spectroscopy. *Bull. Chem. Soc. Ethiop.* **34**, 449–462 (2020).



28. Mengesha, M. H. Chemical composition of teff (*Eragrostis tef*) compared with that of wheat, barley and grain sorghum. *Econ. Bot.* **20**, 268–273 (1966).
29. Baye, K., Mouquet-Rivier, C., Icard-Vernière, C., Picq, C. & Guyot, J.-P. Changes in mineral absorption inhibitors consequent to fermentation of Ethiopian injera: implications for predicted iron bioavailability and bioaccessibility. *Int. J. Food Sci. Technol.* **49**, 174–180 (2014).
30. (ifpri), I. F. P. R. I. & International Food Policy Research Institute (IFPRI). The economics of teff: Exploring Ethiopia's biggest cash crop. Preprint at <https://doi.org/10.2499/9780896292833> (2018).

## **CHAPTER 5: FUTURE DIRECTIONS**

In this work, I presented the teff association panel and evaluated its phenotypic and genetic diversity. I showed that with abundant genetic diversity across nine subpopulations, the panel can be used for genetic discovery. Pairing the large-scale multi-year phenotypic dataset with this re-sequencing data can aid in the selection of lines with specific trait profiles for future teff breeding research.

In the first chapter, I introduced the incredible resilience of millets and highlighted their potential to improve the global agricultural food system. Combined with its C4 physiology, teff has a high water use efficiency that aids in soil moisture preservation resulting in a low rainfall requirements of 300 mm, compared to 500-900 mm for winter wheat and maize (Figure 1; Chapter 1). Although the drought tolerance of teff is well documented, accurate screening with paired physiological data is limited. I conducted a preliminary study with PVC rainout shelters and was able to simulate a climate relevant drought stress for teff in the field. To better understand the resilience of teff and how drought stress effects agronomic and nutritional traits, I suggest an improved drought study, utilizing the diversity of the panel as well as a permanent rainout shelter, to phenotype differences across subpopulations including *E. pilosa* in both well-watered and drought conditions.

The second chapter introduced the panel, emphasizing its genetic diversity and how the resequencing data might be utilized. The genome assembly of *E. pilosa* offers a new tool for comparative genomic analyses. I suggest continued research on the putative sweeps identified via XP-CLR and nucleotide diversity. Since we are still interpreting physical differences between pilosa and the TAP and we only have access to a few pilosa lines, it would be beneficial to work backwards from previously identified domestication genes and compare the synteny of those regions in pilosa and teff.

Additionally, we are working to collaborate with Erich Grotewold as well as another VanBuren lab member, Elliot Braun, to screen white and brown teff seed as well as pilosa for phenolic compounds using high-performance liquid chromatography. This data will functionally validate our hypothesis of seed color's correlation to phenolic compounds. Specifically, we can measure kaempferol concentration and characterize genetic differences across the panel in regions surrounding Et\_4B\_037025, or CYP75B1.

In chapter three we began an in-depth review of the agronomic and morphological differences across the panel. The main finding here was the correlation of lodging susceptibility

and panicle architecture in that more open panicles lodge more consistently. Before harvest, panicles were collected and imaged each field season. We attempted to employ image analysis for a more detailed description of panicle morphological diversity; however, with the architectural differences across the panel paired with the minute width of teff branches, annotation of panicle features was unsuccessful with published protocols like ImageJ, PlantCV or Rhizovision. I suggest a smaller study, with a subset of the panel, to be annotated manually. If these preliminary results elucidate new information not gleaned from panicle architecture score, the protocol can be expanded to the entire panel.

Furthermore, additional collaboration with Dr. Getu Beyene at the Donald Danforth Center would benefit the application of our research. His team developed the teff transformation and CRISPR/Cas9 gene editing protocol and has used it to produce gene edited dwarfing lines. Functional annotation of gene candidates from our marker-associations for lodging and panicle architecture would aid in the confirmation of our hypotheses and improvement of future varietal development.

The fourth chapter provided the first large-scale seed micronutrient screening of teff germplasm. This work could be improved with multi-environment field testing. Perhaps after the work has been published, we can set up a collaborative effort to plant the panel or a subset of the panel at multiple locations across the globe. One factor that was briefly discussed in this chapter is seed size. Smaller seed size is often associated with higher nutritional content. We also have images of teff seed from each season and have been working on analyzing them with ImageJ and PlantCV. This dataset would allow us to test the correlation between seed size and mineral nutrient content and improve our understanding of seed yield across the TAP.

Overall, this work could be dramatically improved with Ethiopian trials. Interacting with growers and consumers is the key to agricultural crop improvement. Continued connection with the Ethiopian Institute of Agricultural Research (EIAR) is essential to ensure farmer preference is prioritized and adoption of new and improved varieties can be achieved. An exciting new collaboration between the Donald Danforth Plant Science Center and the EIAR funded by the Bill and Melinda Gates Foundation will focus on gene editing in teff for lodging resistance. My hope is that one aspect of my research will benefit theirs, and that the funding of this project as well as my own will inspire additional research on this outstanding crop.