Mapping the genomic regions underlying the ionome in the model crop foxtail millet (*Setaria italica*) and the proposal of a probabilistic method for their validation

By

Brianna Leigh Haining

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sarah C. Hake, Chair
Professor Sheng Luan
Professor Céline Pallud

Summer 2020

Mapping the genomic regions underlying the ionome in the model crop foxtail millet (*Setaria italica*) and the proposal of a probabilistic method for their validation

By

Brianna Leigh Haining

Abstract

Mapping the genomic regions underlying the ionome in the model crop foxtail millet (*Setaria italica*) and the proposal of a probabilistic method for their validation

by

Brianna Leigh Haining

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Sarah C. Hake, Chair

Most cultures obtain the majority of their caloric and mineral nutrition from cereals either directly or through livestock; therefore, understanding the genetic architecture underlying the acquisition of ions from the soil in these crops is important for agricultural purposes. The ionome of an organism is intricately interconnected; many transporters and enzymes involved in ion uptake and storage have been implicated in the homeostasis of multiple elements. Thus, an elemental signature of iron deficiency can be detected in an organism through analysis of the entire ionome. The focus of this thesis is on the ionome of two closely related species of grass known as *Setaria viridis* and *S. italica.*

This thesis uses standing diversity and constructed inbred lines to illustrate the ionomic variation present in the *Setaria* species complex, and to determine which regions of the genome are associated with this variation. Chapter 1 explores the effect of iron concentration upon the ionome of a diverse group of *Setaria* landraces and cultivars. Each accession was grown in the soil as well as a hydroponic medium containing three varying concentrations of iron. This experiment aimed to assess the variation in iron accumulation present in the species. Plants were grown until senescence and then harvested and dried. Dry root and shoot weight was assessed, as was yield by weight. Additionally, the uppermost leaf was taken and the concentrations of 20 different elements in the plant were assessed. The unsupervised machine learning algorithm DBSCAN allowed for the identification of two separate ionomically defined groups and three morphologically defined groups. These groupings did not show strong intercorrelation, with the exception of one ionomic group, which did correspond to the African morphology group. The second ionomic group appeared to show a constitutive phosphate deficiency response.

Chapter 2 addresses the need to statistically validate the quantitative trait loci (QTL) identified in association studies such as those completed in Chapter 3 and offers a mathematical solution to this issue along with an open source R module. The Scanning Probabilistic QTL Validator calculates the significance of a QTL of a particular length overlaying any N genes that were previously identified for the trait of interest. The strategy put forth in this chapter takes into account the number of QTL that are identified for a particular model, their lengths, the number of markers used for the initial mapping, and the overall gene distribution in the organism of interest. The SPQV's usage is demonstrated by validating the results of a QTL mapping experiment in the TeoNAM RIL population that was aimed at identifying genes associated with branching.

Chapter 3 details the use of the quantitative trait locus (QTL) mapping to identify regions of the *Setaria* genome associated with ionic homeostasis in the species complex. A RIL population resulting from the interspecific cross between *S. viridis* and *S. italica* was grown in treatments designed to assay the effects of planting density and water availability. As in Chapter 1, the flag leaves of these plants were harvested and subjected to ICP-MS to assay their ionomic content. These phenotypic data were then used to map the regions of the genome that are associated with alterations in the ionome of the species complex. Further mapping was performed using the rotated loadings of principal components analyses which were performed on the phenotypic data resulting from the initial grow-outs. A total of 251 QTL were identified. Multiple concentrated regions of QTL were identified that overlapped with regions previously identified as important for the trait of water use efficiency.

Table of Contents

Figures

Tables

Acknowledgements

This thesis would not exist without the support of many people, and here I take the opportunity to extend my sincerest gratitude to them. First and foremost, I am deeply grateful to my advisor, Professor Sarah Hake. Thanks to Sarah's guidance and unrelenting faith in me, I was able to pursue the field of inquiry that first drove me to graduate school. I would also like to thank the members of my committee, Professors Sheng Luan and Céline Pallud. Their input on my proposed research was vital to the success of this project.

I'd like to express my appreciation of the many people at the Plant Gene Expression Center who have supported and worked with me throughout my PhD. The members of the various labs at the PGEC were always kind and a source of great scientific discussion, particularly Dr Jake Brunkard. I of course owe particular thanks to the members of the Hake Lab, including China Lunde Shaw, Dr Samuel Lieboff, Dr Angus Vajk, Dr Zhaobin Dong, and Dr Annis Richardson. Dr George Chuck in particular was endlessly helpful for my project. Dr Alyssa Anderson was my introduction to both the Hake lab and to graduate school in general. Philosophically, personally, scientifically, and artistically, Alyssa was integral to my advancement, and for that she has my eternal gratitude. My collaborators in the Baxter Lab, including Professor Ivan Baxter, Dr Max Feldman, and Dr Greg Ziegler were vitally important to my introduction to computational biology.

The students who worked with me in the lab were an excellent source of motivation, and allowed me to learn what it is to teach. Ryan Martinez was my first intern, and I thank him for taking that leap with me. Christine Raj and Shannon Prendergast were his intelligent and engaging successors. I am incredibly proud of what my students have already accomplished, and am excited to see where their paths take them.

I would be remiss to write these acknowledgements without including some of the teachers and professors who influenced my trajectory. Mark Emery of Inglemoor High School was the first to uncover my love of biology, and Sue Black was the first to point me in the direction of plant science. Professor Michael Purugganan gave me the opportunity to work in his laboratory as an undergraduate, giving me early experience in plant science that served me well throughout my doctoral studies.

I owe thanks to several of my classmates from my undergraduate career, and to members of the community in which I have found myself. Eric Johnson was integral to my early success in mathematically dense disciplines. Dr Juan Beltran Lacouture has been a bulwark for me throughout graduate school. His unfailing faith in me helped me to cultivate faith in myself. Katy Blumer has been my collaborator, my cheerleader, and, above all else, my friend.

From late night statistics to daily chats during the COVID 19 pandemic, Katy has been a source of pragmatism and determination. Brooke Hopkins and Martin Slosarik have been incredibly supportive; they are the kindest neighbors a person could want. Naomi Dreyer has my gratitude for being an encouraging and engaging housemate and friend. I would also like to thank Mike Ahoy for our many valued conversations, and for giving me the chance to live affordably in the Bay Area.

The department of Plant and Microbial Biology was an overwhelmingly welcoming milieu. Most notable was Rocio Sanchez, who was extremely helpful in terms of both logistical and personal support. I would also like to thank my cohort for their support, friendship, and thought provoking conversation. I would like to extend particular thanks to Gordon Pherribo for our milkshake-heavy tête-à-têtes, Kris Kennedy for his fascinating perspectives, Dylan McClung for being a wonderful housemate and friend, and Grady Pierroz for reminding me to chill out and eat lunch. Martin Alexander was a dear friend, labmate, and advisor, and he drove me to be better in a multitude of ways. I consider myself profoundly lucky to have had the time he spent with us. From the year above, Dr Siwen Deng was fantastically encouraging, a wonderful exercise and writing accomplice.

Finally, I wish to thank my parents and grandparents for their love and encouragement, without whom I would never have enjoyed so many opportunities.

Introduction

While agriculture in wealthy nations has entered the domain of mass market and industrialization, subsistence farming is still practiced in vast swaths of the developing world. In this model of agriculture, the output of a family's farm is primarily used to feed that family. This agricultural context can have profound consequences for human health, as such limited diets contribute to micronutrient malnutrition. Iron deficiency is the most common deficiency, affecting over 2 billion people including many women of childbearing age and children (van der Waals and Laker 2008). The associated costs are immense in terms of both human life and economic output (W.H.O. 2001). Because communities are often constrained to specific crop species by environmental and cultural factors, it is necessary to pursue local crop improvement in addition to larger scale efforts. Natural variation within closely related species provides the means to manipulate agronomically important traits, iron uptake among these, reducing the severity of iron deficiency in subsistence cultures. Staple cereals including rice and wheat have a wide range of variation in iron uptake and spatial allocation among their cultivars; this has been leveraged previously to enhance their iron uptake properties (Römheld and Marschner 1990; Shazhad *et al.* 2014).

Foxtail millet (*Setaria italica*) is a critical subsistence crop in Northern China, India, and parts of Africa (Brink 2006; Dwivedi *et al.* 2012). Additionally, *S. italica* and its wild ancestor, *S. viridis*, are emerging model organisms. *S. italica* is a relatively recent domesticate, and can be considered a subspecies of *Setaria viridis* in spite of their phenotypic differences (Dekker 2004). Indeed, the two species are capable of cross pollination and exhibit a continuous spectrum of morphological traits. In light of this, they are often denoted *Setaria italica* subsp. *viridis* and *Setaria italica* subsp. *italica*. Throughout this thesis, groups of individuals containing both subspecies will be referred to as either *Setaria* or the *Setaria* species complex. Due to its close relationship with economically important crops like maize and wheat (Figure 0-1), small stature, relatively short life span, status as a C4 plant, and sequenced genome, *S. italica* subsp. *viridis* is becoming a preferred subject for genetic studies (Brutnell 2010; Li *et al.* 2011; Bennetzen 2012).

Mechanisms of iron uptake in plants

In initial studies of plant iron acquisition, it appeared that plants undertook the acquisition of iron according to two distinct strategies, dubbed Strategy I and Strategy II. (Römheld and Marschner 1986; Römheld 1987). In strategy I, plants exude H+ in order to reduce the insoluble Fe^{III} to the soluble Fe^{II}, increase root surface reductases that reduce Fe^{III} to Fe^{II}, and, occasionally, release reducing or chelating phenolic compounds into the soil solution (Chaney *et al.* 1972; Römheld *et al.* 1986; Hether *et al.* 1984).

Strategy II plants are characterized by the iron deficiency induced release of Fe^{III} chelating compounds known as phytosiderophores (PS) and the presence of transporters for Fe^{III}-phytosiderophores (Figure 0-2; Takagi *et al.* 1984; Römheld and Marschner 1986). The different strategies were originally assumed to belong to dicots on the one hand (Strategy I) and monocots on the other (Strategy II), and then in later years were proposed to belong to all higher plants except grasses and grasses, respectively (Römheld 1987). Recent evidence points to a much more permeable boundary between the two strategies, as both graminaceous and non graminaceous plants seem to share a broad homology in uptake strategies (Xiong *et al.* 2013; Suzuki *et al.* 2016; Grillet and Schmidt 2019). Rice, specifically, has been proven to exercise both strategies depending upon soil conditions (Ishimaru *et al.* 2006; Walker and Connolly 2008; Wairich *et al.* 2019).

Long distance transport of iron in plants has been the subject of less focus. It is known that in the xylem, Fe^{III} is usually complexed to citrate (Durrett *et al.* 2007). From there, iron is shuttled to the aerial parts of the plant. Iron is not restrained to travel by means of the xylem; older leaves will remobilize and transport iron to newer sinks (such as emerging leaves and seeds) through the phloem (Tsukamoto *et al.* 2008). Iron in the phloem is usually found complexed to nicotianamine (Koike *et al.* 2004). It is thought that the different chaperones involved in the xylem and phloem have to do with the differing pH of these vascular tissues. Once in the leaf, iron is stored in the mitochondria, chloroplast, and vacuole. In these locations, iron is typically found complexed to either ferritin or phytate (Figure 0-2; Grillet *et al.* 2014).

Setaria species are in the Poaceae, and so utilize the PS based method Strategy II. This pathway centers around the exudation of mugineic acid-family phytosiderophores (MAs), which are small metal chelators (Hell and Stephan 2003). These molecules are synthesized from a starting material of L-methionine, which is converted to nicotianamine and then to various mugineic acids. MAs are synthesized and stored in vesicles in the epidermal cells of the root (Negishi *et al.* 2002). The MA is then transported to the soil through the cooperation of *yellow stripe like* 5 (YSL5), a vesicular transporter, and *transporter of mugineic acid* (TOM), a transporter that is localized to the plasma membrane (Zheng *et al.* 2011). Once in the soil, the MAs chelate ferric iron and are then taken up again by the plant through YS and YS-like (YSL) channels (Kobayashi and Nishizawa 2012; Curie *et al.* 2001). In root epidermal cells, Fe^{III} is loaded into the xylem and transported towards the shoots (Figure 0-2). Due to free iron's tendency to generate reactive oxygen species, iron is chaperoned throughout the plant by either nicotianamine, MAs, or citrate (Morrissey *et al.* 2009, Kobayashi and Nishizawa 2012). *Setaria's* genome also contains several members of the NRAMP family of genes, which have been shown to mediate the uptake of Fe^{II} and other metal ions in other species, including barley (Thomine and Schroeder 2013; Wu *et al.* 2016). The presence of NRAMP transporters suggests the possibility that *Setaria* functions at the intersection of the two strategies for iron uptake.

The Ionome: Ionic homeostasis as a composite trait

Iron acquisition does not occur in a vacuum; ionic homeostasis is intricately connected (Figure 0-3; Shakoor *et al.* 2016). The concentration of iron plays a role in influencing the concentration of other ions through its control over plant height. Ions such as zinc, copper, and manganese share enough properties with iron to be chelated by PS and taken up by the same transporters (von Wirén *et al.* 1996; Römheld 1991), though the transporters do seem to be exclusively regulated by iron status. In rice, iron deficiency-associated upregulation of the iron transporters OsIRT1 and OsIRT2 contributes to cadmium uptake (Nakanishi *et al.* 2006). Ion interactions with the soil solution also impact the bioavailability of other, complementary ionic species. If there are multiple ions held very tightly to a given colloid within the soil, they may work to 'shield' nearby ions from plant roots, thereby decreasing the availability of an ion that is typically fairly accessible (Brady and Weil 2016). The interconnection of ions in the grasses is not entirely dependent upon iron; magnesium and cadmium compete for translocation in barley (Kudo *et al.* 2015), calcium impacts concentrations of manganese, potassium, strontium, rubidium, and zinc (Baxter 2008).

It is clear, then, that the relative concentration of one ion in the soil can impact the ultimate concentration of *different* ions in the plant, and that any attempt to understand the genetics underlying the homeostasis of one ion must necessarily address the complex dynamics of the complete ionic milieu. The interconnection of the concentration of various ions has led researchers to the concept of the *ionome* of an organism; that is, its mineral nutrient and trace element composition (Lahner *et al.* 2003). The ionome is essentially the inorganic component of an organic system. The study of the ionome is referred to as ionomics. This process relies on the quantitative measurement of the ionic composition of a living organism as it varies with genetic background, developmental timepoints, and physiological conditions (Salt 2004). Ionomics relies on the simultaneous sampling of the entire ionome; this sampling is often conducted with the use of Inductively Coupled Plasma-Mass Spectroscopy (ICP MS; Baxter *et al.* 2008).

Conclusion

In light of both the current agricultural system and the impending upheaval associated with shifts in regional and global climates, we must act quickly to understand and optimize the crops on which we depend. The complete elucidation of the genetic basis of ionic homeostasis in crop species is fundamental to the effort to increase their overall yield and their nutritional value. Species like *S. italica* play an important role in this dynamic, straddling the boundary between crop and model organism.

Figure 0-1

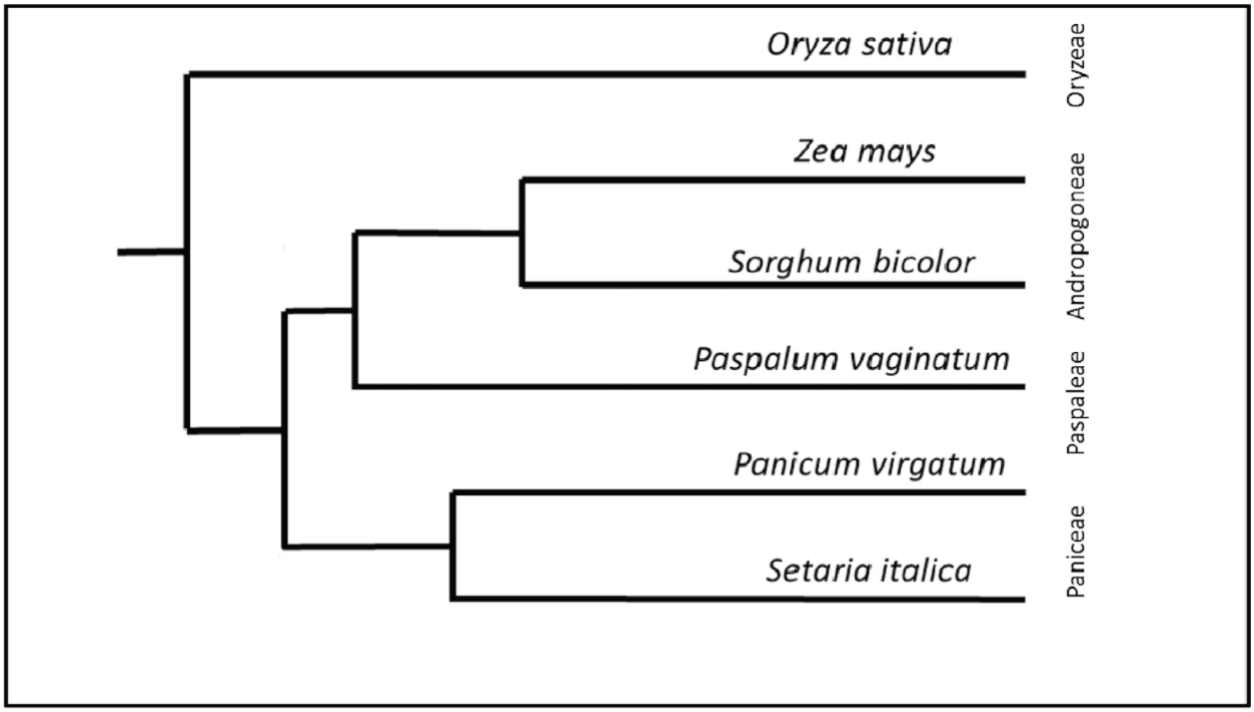


Figure 0-1. The relationship of *S. italica* to other grass species
Setaria is a C4 plant that is closely related to maize, rice, and sorghum.

Figure 0-2

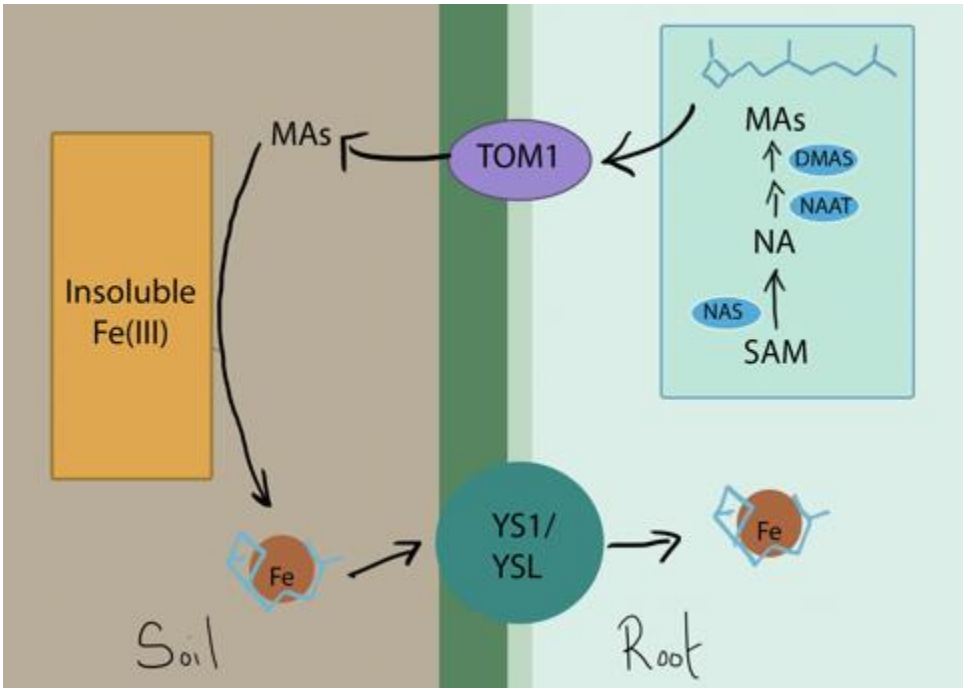


Figure 0-2. Strategy II iron uptake

Modified from López-Arredondo *et al.* 2013, Figure 4. Plants secrete mugineic acid (MA) family phytosiderophores into the rhizosphere, which then chelate insoluble Fe^{III} from soil particles. These chelates are then taken up by way of YS family transporters.

Figure 0-3

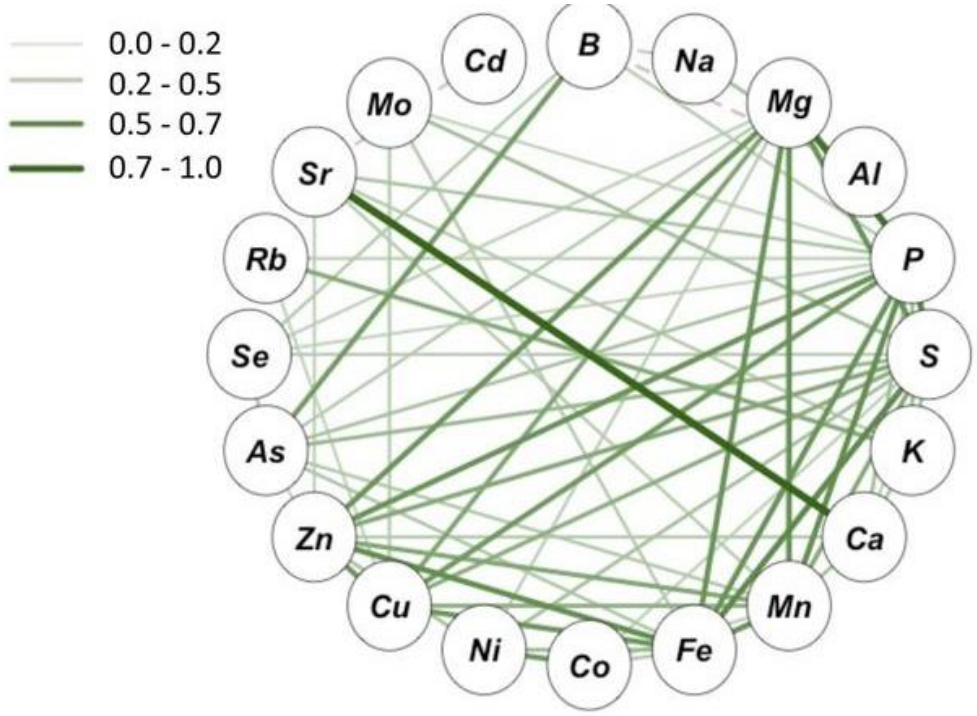


Figure 0-3. The interconnectedness of the ionome in a Strategy II plant

Green lines indicate positive correlation values; red lines indicate negative values. Line thickness indicates the strength of the correlation. The assayed plants were from the Sorghum Association Panel (Casa *et al.* 2008). Modified from Shakoor *et al.* 2016.

Section One. Assessing variation in the ionome of foxtail millet

PREFACE

In light of the recent rise of the Setaria species as model organisms, their inherent diversity has become a subject of interest about which little is known. To elucidate the degree of variation present in the Setaria species complex, a diverse collection of accessions selected from the United States Department of Agriculture's Germplasm Resources Information Network was grown in both soil and hydroponic conditions. The ionic content of these plants was then assessed using mass spectrometry, and the resulting data was analyzed in order to quantify the variation present in the tested accessions.

Chapter 1. Exploring diversity in *S. italica's* ionic homeostasis using machine learning and hydroponics

INTRODUCTION

Natural variation in the homeostasis of iron in *S. italica* remains largely unexplored. Ravindran *et al.* (1991) examined four varietals of *S. italica* for their available proximate and mineral nutrition; one of these was obtained from a seed-breeding station in Sri Lanka, while the other three were obtained from a station in India. These varieties varied widely in their iron content, with the Sri Lankan variety containing nearly twice as much iron as two of the Indian varieties. RFLP analysis of the ribosomal DNA intergenic spacer region in *S. italica* indicates that there are at least two types of rDNA in the species, with at least ten subtypes in total. The variation in rDNA sequence appears to be associated with geopolitical boundaries (Eda *et al.* 2013). Three races of foxtail millet have been defined in terms of morphology, though some have proposed a fourth race known as *nana*. The species has been further divided into ten sub races (Table 1-1; Prasada Rao 1987; Dwivedi *et al.* 2012). Given the identified genetic and morphological diversity of this species, it is likely that variation in ionic content generally, and iron content specifically, exists in *Setaria*. However, subsequent work aimed at defining the breadth of that variation is sparse or nonexistent. A pilot study was therefore devised to determine if sufficient variation existed in the *Setaria* species complex for improvement of the species.

Potential contributors to variation in iron homeostasis in Setaria

Soil type at origin: Because iron is central to the growth, development, and ultimate reproductive capacity of plants, the different availabilities of this nutrient in different soils provides a strongly variable selective pressure. Though *S. italica* appears to have been domesticated from *S. viridis* in Gansu Province, Northwestern China in around 5900 BP (Barton *et al.* 2009), it has since been cultivated in many diverse regions, including 'Southeast Europe, South and Central China, the Far East, the Americas, Australia', and parts of Africa (de Wet *et al.* 1979). Given this wide pattern of cultivation, it is likely that *S. italica* has been cultivated in many different soil types. Indeed, the USDA's Germplasm Resources Information Network (GRIN) contains 772 *S. italica* accessions at time of writing. The documentation for these accessions contains information such as the exact coordinates of collection, which indicate the diversity of soils in which *S. italica* has been cultivated during the last century.

The diversity of soil characteristics is commonly defined by the USDA's 12 order system (Brady and Weil 2016; Table 2-1). These orders differ in a wide variety of properties, many of which are important to ionic homeostasis in the plants that live upon the soil. Among these characteristics are the percentage of organic matter, mean annual soil temperature, base saturation, pH and cation exchange capacity (CEC). CEC is defined as the total sum of exchangeable cations (such as Fe) that are adsorbable by the soil.

CEC is largely dependent upon pH: the number of exchangeable cations increases in the soil solution as pH increases (Brady and Weil 2016). Phytosiderophore (PS) uptake in the grasses also depends upon pH, and therefore on soil type, with lower pH contributing to increased PS uptake (Schaaf *et al.* 2004). There is evidence that soil type influences PS exudation as well: the Tamaro wheat cultivar appears to exude different levels of PS in different soils (Oburger *et al.* 2014). This is consistent with the differential induction of PS in media with different iron availabilities (Itanna and Coulman 2003; Gries *et al.* 1998; Kabir *et al.* 2016).

Biotic milieu at origin: Plants working to obtain mineral nutrients not only face the obstacle of extracting insoluble iron from the soil, they also face significant competition from organisms in the rhizosphere for these resources. The appropriate transporters can give strategy I plants the advantage by giving them the ability to take up Fe^{III}-PS complexes originally secreted from nearby strategy II plants. For example, iron efficient oat (*Avena byzantina* C. Koch.) produces PS that are taken up by iron inefficient muskmelon (*Cucumis melo* L.) and T3238fer tomato *(Lycopersicum esculentum* Mill.) when they are hydroponically co-cultivated (Camp *et al.* 1987; Hopkins *et al.* 1992). If a strategy I plant is unable to take advantage of the Fe^{III}-PS complexes in this manner, they are at a comparative disadvantage: soybeans which were unable to take up these compounds in the same studies suffered from increased iron-deficiency chlorosis as compared to plants grown alone.

Another source of biotic complexity is found in the microbiome. Microbes take up iron and other essential nutrients using siderophores, molecules that are similar in structure to PS and which are taken up using similar transporters; the negatively charged groups of microbial cell walls also adsorb cations like Fe^{II}, thus removing them from the soil solution (Rajendran *et al.* 2003). While there is general interchange between plant and microbial siderophores (Bar-Ness *et al.* 1992; Jurkevitch *et al.* 1993; Radzki *et al.* 2013), PS also serve as attractive carbon sources for microbes in the soil (Marschner and Crowley 1998; Von Wiren *et al.* 1993). Microbial metabolisms can act to decrease the soil pH, particularly in Fe^{II} poor environments (Lauber *et al.* 2009); in conditions where the necessary cations are freely available, however, the microbiome can act to increase the pH of the soil (Ratzke and Gore, 2018). The impact of the soil microbiome on iron uptake strategies that do not depend on PS can therefore be acute, as Fe^{II} availability depends heavily upon soil pH.

Degree of domestication and morphotype: Strains in the *Setaria* complex differ broadly in key domestication traits. A non-shattering panicle, limited tillering, and appropriately timed seed germination are all traits that contribute to a plant's domestication syndrome, and none are fixed in agriculturally utilized lines of *Setaria*. Indeed, domesticated *Setaria* has been hypothesized to consist of three races: *moharia*, *maxima*, and *indica*. These races are defined by morphology. *Moharia* lines are highly tillered and produce many small, upright panicles, much like the wild *S. italica* subsp. *viridis*.

Those lines classified as *maxima* tend to produce only one or two large, pendulous panicles, and do not generally tiller. *Indica* lines appear to be an intermediate between the other two races in terms of tiller number and panicle habit (Prasada Rao *et al.* 1987; Table 1-1). While *maxima* lines conform best to the expected domestication syndrome, there exist elite cultivars that are considered to be of the *moharian* race. The tripartite morphological division is not universally accepted: Li *et al.* (1996) apply a fourth designation, *nana*, to the cultivars that most closely match the phenotype of *S. italica* subsp. *viridis*. In their designatory scheme, *moharia* has more domesticated characteristics than does *nana*, with relatively fewer tillers and shorter culms.

There is a strong correlation between morphotype and location of cultivation: *moharian* lines are grown in Europe and southwest Asia, *maxima* in transcaucasian Russia and eastern Asia, and *indica* lines in India and southeast Asia. Li *et al.* claim that *nana* is found in Lebanon, Iran, and Afghanistan. It is apparent that, though race, degree of domestication, and geographic origin are inextricably intertwined, these factors do operate independently of each other to an extent, largely due to intensive breeding and cultivation in the United States, China, and Europe. Because *maxima* is the most suitable for modern cultivation practices, it is likely that the cultivation of this race has been preferred in regions where intensive breeding of *S. italica* has been carried out. The regional associations discussed in this paragraph were largely replicated in a RFLP analysis (Eda *et al.* 2013) of the ribosomal DNA intergenic spacer sequence, with the exception of the *nana* race, which was not clearly present in their analysis. Additionally, the RFLP analysis suggests that there is recent shared parentage between most African and Indian lines.

The known variation in iron content in *S. italica* combined with the likelihood that one or more of the factors described above suggest that there exists unexplored variation for this phenotype in the *Setaria* species complex. A study was therefore devised to determine if sufficient variation existed for improvement of the iron content in the species. A hydroponic system was developed to assay the impact of varying concentrations of iron on the morphology and ionome of numerous accessions of *S. italica* collected from around the globe. All three races were represented in these accessions. A hydroponic system was used to assay the ionic content, as these systems are excellent tools in plant nutritional studies (Jones Jr. 1982). As a complement to the hydroponic system, the accessions were also grown in soil and analyzed in the same manner as the hydroponic system. Unsupervised machine learning methods were leveraged to group morphological and ionomic observations. These data revealed three distinct morphological groups segregated along geopolitical boundaries and two distinct ionomic groups containing accessions of mixed geopolitical origins, thus indicating that morphology does not define the *S. italica* ionome. Additionally, valuable data were collected on the influence of morphotype on robustness and response to hydroponic growth. The insights gained through this study demonstrate that sufficient natural variation in Fe homeostasis exists in the *S. italica* pan-genome for use in biofortification.

MATERIALS AND METHODS

Selection of the diversity panel

Sixteen lines of *Setaria italica* were selected from the Germplasm Resources Information Network (GRIN), a germplasm database that is maintained by the United States Department of Agriculture's Agricultural Research Service. The chosen lines are detailed in Table 2-2. The lines were selected for diversity primarily based on their soil type at their approximate location of collection or origin (when known). Eight out of the 12 USDA-recognized soil orders were represented in the selected lines. The unrepresented orders are gelisols (comprising 13% of continental land area), spodosols (4%), histosols (2.5%), and vertisols (2%). Though combined these soil orders cover 21.5% of the earth's surface, gelisols are defined by the presence of permafrost; the other orders that are not included are also typically non-arable due to the cold environment associated with the formation of those soils. Annotation concerning the exact location of an accession's collection was sparse in the GRIN database; some accessions were only identified by their country of origin. Kenya and Afghanistan have extremely complex soil landscapes (Jones *et al.* 2013; Food and Agriculture Organization of the United Nations, 1973), and *S. italica* was originally domesticated in China (Lu *et al.* 2009). These countries were therefore preferentially selected.

Degree of cultivation was also taken into account when establishing the panel; the majority of tested accessions were landraces, as crops at this level of domestication tend to be more genetically diverse than elite cultivars. Single accessions of *Setaria viridis*, *Setaria parviflora*, *Setaria faberi*, *Setaria macrostachya*, and *Setaria sphacelata* were also included in the panel. The *S. faberi* accession was a cultivar, and the *S. viridis* line was the lab strain A10. The remaining accessions were originally collected from the wild.

Nutrient Medium Composition

The nutrient medium used in all hydroponic experiments was based on a modified MS solution. The 1x solution consisted of 2.0 mM $Ca(NO_3)_2$, 0.5 mM $MgSO_4$, 0.1 mM KCl, 10 μM H_3BO_3, 5μM $MnCl_2$, 5 μM $ZnCl_2$, 2 μM $CuCl_2$, 1 μM Na_2MoO_4, 300 μM KH_2PO_4 and 0.15 mM FeEDTA (Ceasar *et al.* 2014). FeEDTA was chosen as the iron supply for several reasons. Firstly, this iron chelate has been used in several studies aimed at understanding iron homeostasis in grasses (Santana *et al.* 2014; Li *et al.* 2019; Krohling *et al.* 2016). Because phytosiderophores work in the soil solution to chelate iron from colloidal particles, it is important that the ion chelate has a binding constant that produces an environment that is similar to that created by the presence of Fe^{III}-colloid complexes. This means that the binding constant must be high enough that the complex does not dissociate immediately in solution, but low enough that the Fe^{III} can be chelated by a phytosiderophore, both of which are satisfied by Fe^{III}EDTA.

Two other iron media were also used: a solution containing 1 mM FeEDTA and a solution containing 0.06 mM FeEDTA. The high iron (+Fe) concentration was selected based on Santana *et al.* 2014, which illustrated that leaf bronzing, which is a symptom of iron toxicity, can be visible at 1mM FeEDTA in *Paspalum urvillei*, a grass that can tolerate high iron soils. In the same study, *Setaria parviflora*, a hyperaccumulating species, showed bronzing at 2 mM FeEDTA. The intent of this experiment was to supply the plants with a stressful concentration of iron without killing them. Because most species are, by definition, neither hyperaccumulators nor hypertolerant of excess iron, the highest concentration of iron that did not visibly stress *S. parviflora* was used. The low iron (-Fe) concentration was selected based on the interactions between iron and potassium in maize that were elucidated in Celik *et al.* 2010. When exposed to conditions of potassium ranging from 1-8 mM and 0.03 - 0.12 mM FeEDDHA in an otherwise standard nutrient medium, the dry weights of maize roots and shoots reached their peaks at 4-6 mM potassium and 0.12 mM FeEDDHA. At 0.06 mM FeEDDHA, the dry root weight was decreased significantly in all potassium regimes, and the dry shoot weight was decreased significantly in all but the 8 mM potassium treatment. This concentration of ferric chelate was therefore used in an attempt to produce chlorotic plants that still set seed. All media types were produced in 20x solutions, which were adjusted to a pH of 6 using NaOH and glacial acetic acid where necessary.

Germination procedure

Several methods of germination were assayed in hopes of achieving simultaneous germination. Each method was performed using unstratified seed and seed that had been held at 4℃ for a minimum of 5 days prior to surface sterilization. Seeds were then surface sterilized in a solution of 10% bleach for 2 minutes before germination. In method one, the sterilized seeds were placed on a damp paper towel, which was then enclosed in a petri dish and left in the dark to allow for germination. Water was added to the paper towels as they dried. In method two, seeds were placed in petri dishes and soaked in a karrakin containing solution (Wright's Hickory Liquid Smoke) in the dark. Methods one and two were each performed at room temperature (~25℃) as well as at 30℃ . In method three, seeds were distributed on 1% agar plates and left in the dark at room temperature.

Method three was successful in producing uniform germination across all assayed accessions and was therefore used in all subsequent experiments.

Cultivation in the Hydroponic System

Germinated seeds were rooted in a fully hydrated mixture containing 70% perlite and 30% vermiculite when their radicles had reached approximately 1 cm in length. Seedlings were blocked by experimental treatment. Each block was supplied with 50% nutrient medium the day following planting through bottom watering. Every three days thereafter, seedlings were supplied with the appropriate nutrient medium at full strength.

When the seedlings had reached the four or five leaf stage, they were transferred to a hydroponic system (Figure 1-1). All *S. italica* varieties were placed in the system on the same day; *S. macrostachya* and *S. sphacelata* reached the appropriate stage and were placed in the system two weeks after the other varieties. Each plant had a single neighbor of the same cultivar within the same hydroponic unit (HU). HU containing plants of the same genotype were placed in physical proximity in an attempt to control for the effects of microenvironments. Once in the system, plants were supplied with fresh nutrient medium every three days. Leaf number, shoot height, tiller number, and panicle number were assessed on those days.

Within the first ten days, 24/90 plants had died; by 13 days post hydroponic placement (DPHP), that number had increased to 40/90. Examination of the system revealed that the most likely cause of this phenomenon was the physical position of a plant's HU. Plants placed in the two rows closest to the western windows had a much greater likelihood of dying than their neighbors. These plants had HU that were hot to the touch by the late afternoon. Several steps were taken in response to this issue. First, HU positions were altered every three days. All units were transferred to positions previously occupied by another unit in their treatment group in order to control for physical position and to ensure that all plants occupied the same positions in the course of the experiment. Secondly, all plants were separated into individual HU on 8/24/2016 in order to remove the discrepancy in treatment between plants with living neighbors and plants without.

Plants were harvested when they had set seed. Due to the broad range of domestication traits in the cultivars assessed, the date of collection varied widely; while White Wonder produced only two panicles per plant, some produced upwards of 100. In spite of the short life span of most *Setaria* species in soil, many of the cultivars assayed in the hydroponic system appeared to thrive in the context of the hydroponic system, with new tillers and panicles emerging until the date of harvest, six months post germination. These indeterminate plants were harvested simultaneously. Plants were dried for one week before collection of final morphological data, including dry root and shoot weights.
Material was collected from the flag leaf on the main axis of the plant and sent to the Baxter lab at the Donald Danforth Plant Science Center (DDPSC) for mass spectroscopic ionomic analysis (ICP-MS).

The Baxter lab has a documented, standardized pipeline for elemental analysis (Ziegler *et al.* 2013). Flag leaf samples collected from each plant were dried, weighed, and digested in 2.5 mL of concentrated nitric acid (AR Select Grade, VWR International, LLC) with internal standard added (20 µg L–1 Indium 200) (Aristar Plus, BDH Chemicals). They were then diluted to 10 mL using ultra pure water (18.2 MΩ water) from a Milli-Q system (Millipore).

Concentrations of the elements B, Na, Mg, Al, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd were measured in each sample using an Elan 6000 DRC-e mass spectrometer (Perkin-Elmer SCIEX) connected to a PFA microflow nebulizer (Elemental Scientific) and Apex HF desolvator (Elemental Scientific). Nitrogen concentration was not considered, as this technique does not allow for its measurement. A calibration curve was produced before each run by analyzing six dilutions of a stock solution that was produced by mixing multiple single element standards (Ultra Scientific).To reduce interference due to the presence of polyatomic and double charged species, the lens voltage and nebulizer gas flow rate of the ICP-MS were optimized for maximum Indium signal intensity and low CeO+/Ce+ (<0.008) and Ba++/Ba+ (<0.1) ratios. Machine drift within and between runs was corrected for by the inclusion of a control solution after every ten samples. This control sample was the result of mixing the remaining samples from the second dilution. The control therefore reflected any drift in the sample matrix. This same control was used in each run, so that inter-run variation could be corrected.

Cultivation of Soil Grown Plants for comparison

Due to their liquid, axenic media and the lack of the support and resistance offered by soil, hydroponically based experiments can differ drastically from soil-based experiments. Moreover, when compared to soils, hydroponic systems offer well-balanced nutrition to plants while simultaneously providing very little oxygen. In light of these limitations, the cultivars that were grown hydroponically were also grown in soil (Sunshine Mix #2). Using the same criteria as in the hydroponic experiment, plants were harvested at maturity and their shoots dried for one week. Samples from the flag leaves on the main axis of the plant were collected and sent to the DDPSC. They were subsequently analyzed by ICP-MS as in the hydroponic experiment.

Computational Analyses

Due to the high rates of death in the hydroponic system, a generalized linear model (GLM) was fitted to determine the contributing factors. The Akaike Information Criterion estimator was calculated for models including factors such as treatment, genotype, location, and the interactions between these factors.

Analyses of morphological data: Given that *Setaria* is a relatively recent domesticate, it was unclear if cultivar and landrace designations would indicate meaningful distinctions. Several alternative groupings were therefore assessed, including soil type at origin, treatment, race, domestication level, and region of origin. Because the information about accessions in the GRIN database is often incomplete, the morphological data were also grouped using the density-based spatial clustering of applications with noise (DBSCAN; Ester *et al.* 1996) algorithm, with epsilon set to 0.15 and the minimum cluster size set to 8.

This particular algorithm makes no assumptions about the shape and size of individual clusters, or how many clusters exist in the data. The DBSCAN algorithm was run using the data for height, tiller number, total panicle weight, average panicle weight, average panicle length, and average leaf number per tiller. These data were scaled using the scikit-learn (version 0.22.2; Pedregosa *et al.* 2011) package's StandardScaler function, and then converted to a squareform distance matrix using the squareform function from the package scipy 0.14.0 (Virtanen *et al.* 2020). The metric used for this process was 'correlation'. To assess whether the groups assigned by DBSCAN were more informative than random choice and to determine the features that were most informative for this grouping, the clusters defined by DBSCAN were used as the target variable for a random forest regression with 200 estimators. K-fold cross validation with K = 3 was used to prevent overfitting.

The morphology of hydroponically grown plants was compared with that of soil grown plants using PCA. Calculations were performed both with and without wild species using the native prcomp() function. The traits analysed included the traits used in the DBSCAN analysis. Additional PCA were performed separately on the soil- and hydroponically grown plant data.

The data were also subjected to a second method of high dimensional data visualization known as Potential of Heat-diffusion for Affinity-based Transition Embedding, or PHATE (Moon *et al.* 2017). The morphological data for the two separate experiments was visualized separately and alone.

Analyses of ionomic data: PCA was performed on these data as described above. The traits assayed included the contents of the elements Mg, P, S, K, Ca, Fe, Mn, Co, Ni, Cu, Zn, Se, Rb, Sr, Mo, and Cd. Additionally, PHATE was performed on the reduced ionomic dataset, and ionomic groups were determined using DBSCAN with an epsilon of 0.46 and a minimum cluster size of 11. A random forest regression with 3-fold cross validation was performed as described above.

RESULTS

Morphology and location in the HS affected survival

DBSCAN (morphology cluster), DBSCAN (ionomic cluster), and physical location were assessed as possible contributing factors to the widespread death of HS plants, as were the interactions between these factors and treatment. The morphological DBSCAN cluster had a statistically significant effect on survival ($p < 0.001$), as did location ($p < 0.01$); no other factors or interactions appeared to modulate plant survival. No morphological attributes were significantly associated with the ionomic DBSCAN clustering.

Clustering analysis revealed three distinct morphological groups

The results of the DBSCAN clustering revealed three morphological groups centered in Africa, Russia and China (SR, for Sino-Russian), and Afghanistan/Lebanon (AL) (Figure 1-2). When the 'group' variable was assigned at random, the assignment was accurate 25.01% of the time. When the frequency of each group's occurrence was taken into account, the accuracy rose to 27.03%. Random forest regression assigned groups with an accuracy of 83.93%. The features that had the most importance for the clustering are reported in Table 2-3. The most important among these included panicle dimensions and plant height.

Tiller and leaf number allow for early morphological group differentiation

Group means for tiller number, leaf number, and leaves per tiller did not differ between treatments. The low iron group did differ significantly from the high iron group in terms of height on days 25, 37, 43, and 100, with the high iron plants being significantly shorter than their low iron counterparts (Figure 1-3C).

Comparison between races was complicated by the large fraction of lines whose race was not noted in the ARS GRIN database. There nonetheless appeared to be consistent differences between *moharia* and *maxima* plants in terms of height, with a consistent statistically significant difference beginning at 31 DPHP (Figure 1-3B). The two races also differed in leaf and tiller number starting at 22 and 16 DPHP, respectively (Figure 1-4B; Figure 1-5B). Leaf number per tiller allowed for the differentiation of *moharia* and *maxima* from 16 - 43 DPHP, but not at the final measurement (Figure 1-6B).

Different pairs of the DBSCAN generated groupings were distinguishable by height, tiller number, and leaf number. African lines were differentiable from SR plants via leaf number measurements from 7 DPHP onwards (Figure 1-4A), and were significantly different from AL plants in terms of height at 100 DPHP (Figure 1-3A). AL plants were distinct from SR plants in terms of height at 100 DPHP, and in terms of tiller number from 10 DPHP onwards (Figure 1-5A). SR plants had a high leaf number per tiller ratio for the first 43 days of the experiment, and were distinguishable from the AL varieties by this metric from 28 - 43 DPHP (Figure 1-6A).

The control and low iron treatments in the hydroponic system promoted branching

The final morphological data was visualized using PHATE and PCA. When the soil plants were examined alone, the first principle component explained 62.5% of the variance in the data, and the second accounted for a further 21.7% of the variance (Figure 1-7).

The three morphological groups defined by DBSCAN were clearly visible in both visualization methods (Figure 1-7). All three groups separated along the first PC, and AL and Africa were distinguished from SR along the second. The PHATE visualization illustrates that the morphological DBSCAN groups are clearly distinct within the data. With the exception of AL, however, these groups are not obvious in the morphological data of the hydroponic system (Figure 1-8A). It appears that the majority of the members of the African and SR groups did not survive the HS. Instead, the HS morphology appears to be defined by treatment, with the plants treated with the high iron regime clustering in PHATE visualizations and near the origin in PCA analyses (Figure 1-8C; Figure 1-9). When the soil and HS grown plants were compared in PCA, the high iron regime plants clustered closest to those grown in the soil (Figure 1-9A), likely due to the fact that these groups were both short and had a low level of branching in comparison to the plants grown in the control and low iron conditions. Unlike the soil grown plants, plants in the high iron treatment were also observed to have signs of bronzing. The low iron treatment did not appear to induce chlorosis.

Clustering analysis of mass spectroscopy data revealed two groups

The ionomes of the soil grown plants were compared using PHATE. When the ionomic data were colored according to the groups designated by applying DBSCAN to the morphological data, the African group remained clearly defined (Figure 1-10). The AL and SR groups, however, did not appear to have the same similarity in their ionomic properties as they did in their morphologies. In light of this, the DBSCAN algorithm was used to determine the groups within the ionome. This yielded two groups (Figure 1-10B), one of which (Group 1) corresponded well to the African morphology group, though Group 1 did include an Indian cultivar (464157) that the African morphology group did not. The second ionomic group (Group 2) was composed of the A10 strain of *Setaria viridis* and the elite cultivars White Wonder and Poltavskaja. In Group 2 as compared to Group 1, there is a significantly higher level of As, Zn, P, Cu, and Fe, and significantly lower levels of Co. B did not differ between groups (Figure 1-11).

Because the morphological clustering did not appear to correspond well to that of the ionome (Figure 1-10), the ionomic DBSCAN clustering was assessed using a random forest with the same parameters as described for the morphological clustering. Random selection of the ‘group’ variable had a 33.32% success rate; when group frequency was accounted for, this rate rose to 35.15%. Random forest regression assigned groups with an accuracy of 82.00%. The most informative elements for this clustering were Fe, Mo and Mg (Table 2-4).

The ionomes of the HS plants were similarly analysed using PCA and PHATE. When assessed alone, the HS plants did not exhibit a clear pattern of clustering, which is likely due to the relatively small number of plants that survived the HS.

When the ionomes of all plants were considered simultaneously, a stark division between the soil grown plants and the HS grown plants is clear (Figure 1-12C). Clustering by HS treatment regime is more evident in this context, with clear distinctions between those plants grown in the high iron regime and their counterparts in the low and control iron regimes.

DISCUSSION

Constitutive phosphate deficiency response defines the ionome in Setaria italica

K fold cross-validation shows that Fe was the most important feature for the designation of ionomic groups using DBSCAN. The importance of Fe in determining these groups suggests that there is sufficient variation in the homeostasis of Fe to differentiate between groups. Additionally, Fe homeostasis is clearly associated with genotype in *Setaria italica*: Group 1 consists of all assayed African accessions and one Indian accession, and Group 2 consists of two elite cultivars and the lab strain of *Setaria viridis* known as A10.

Though the accessions in Group 2 are not morphologically similar, they have been cultivated in similar environments for the past two decades (Wang *et al.* 1998); it is possible that the consistent supply of fertilizer that both lab specimens and elite varieties enjoy has shaped their ionome. A similar phenomenon has occurred in *Arabidopsis thaliana*: though this species only became a common model organism in the 1980s (Meyerowitz 2001), by 1997 both the *Col-0* and *Ler* genotypes had lost secondary seed dormancy (Van der Schaar *et al.* 1997). This sort of unconscious domestication is common, having occurred in fruit flies, bacteria, and the green alga *Chlamydomonas*, among others (Driks and Eichenberger 2016; Bell 2012).

Group 2 accumulated more than twice the amount of iron as did Group 1, which may be a result of the prolonged, heavy use of fertilizer on the accessions in Group 2. Fertilizers containing high levels of phosphorus are applied regularly to both lab strains and elite varieties cultivated in the United States and Europe, with application levels since 1980 ranging from 0.5 grams of phosphorus per meter squared per year to triple that amount (Lu and Tian 2017). In contrast, phosphate fertilizer use in Africa has held steady at less than 0.5 g $P/m_2/yr$ in the same time period. The presence of phosphate has a well-documented impact on the availability of both Fe and P to graminaceous plants, as they interact to form insoluble complexes in the soil (Lindsay and De Ment 1961). It is possible that the ability to take up iron under high phosphorus conditions has been unintentionally bred into the accessions in Group 2.

There is a known multi elemental signature of phosphate deficiency in strategy I plants (Baxter *et al.* 2008; Heuwinkel *et al.* 1992). Aside from a drop in phosphorus content, *A. thaliana* also displays an increase in Mo, B, Zn, As, and Fe, and a decrease in Co and Cu. Group 2 plants follow this pattern when compared with Group 1 plants with the exception of their P and Cu content, both of which were relatively increased in Group 2. Copper is also mobilized by phytosiderophores, but is not mobilized by the pH modulation that is characteristic of the strategy I iron uptake pathway (Printz *et al.* 2016). If these plants have indeed upregulated their iron uptake pathways in response to excess phosphate, then this increase of copper is likely due to an excess of phytosiderophores in the environment. It seems, therefore, that the plants in Group 2 may have been bred to have a constitutive phosphate deficiency response regardless of their actual phosphate contents.

Morphological group in S. italica *contributed to survival in the hydroponic system*

The root system architecture of grasses has been shown to be controlled by the same genetic network that controls tillering (Gaudin *et al.* 2014; Khasanova *et al.* 2019), and a heavily branched root system contributes to survival in hypoxic environments (Koevoets *et al.* 2016; Jacobs *et al.* 1997). This is likely why the HS appears to have selected against plants in the morphological groups with relatively few tillers (African and SR). Interestingly, a plant's ionomic grouping did not appear to affect survival in the HS. The disjunction of groups as defined by morphology and the ionome suggests that the morphology of these plants has little effect on the ionome, and vice versa.

In combination, the differential survival rates in the HS and the lack of overlap in clusters defined by the ionomic and morphological datasets suggest that while an increase in branching does improve *S. italica*'s survival in hypoxic environments, it does not appear to affect ionic homeostasis.

The hydroponic system perturbed the ionome and morphology of S. italica

Many of the plants treated with high iron also had low seed set due to a lack of panicles or the presence of immature panicles. In combination with the visible bronzing on these plants, this finding suggests that the HS was successful in inducing iron toxicity in many of the plants in the high iron condition. Iron deficiency, however, was not visible in the plants treated with low iron media.

Though the different levels of iron exposure did not dramatically perturb morphological grouping by region, the morphological groups defined by DBSCAN did not correspond to the ionomic dataset. While variation in ionic homeostasis does appear to be loosely associated with region, the treatment regime to which the plants were exposed was more obviously represented in the clustering of the data.

Interestingly, while the plants exposed to the high iron treatment were morphologically the most similar to the soil grown plants, they were the most ionomically *dissimilar* to the soil grown plants. Given that the plants in the low and control iron treatments were given relatively balanced nutrient media, the differences in morphology between the soil grown plants and their more branched counterparts in the control and low iron treatments are primarily due to the amount of nutrients available to the plant. Without the high iron treatment's excess of iron to interfere with the uptake of other ions, typical ionomic homeostasis was largely maintained in the low iron and control treatments.

Iron treatment alone did not significantly impact plant form, but the overall growth environment did dramatically impact the morphology of the plants. When the morphology of all plants was compared on the basis of treatment using both PCA and PHATE, the plants treated with excess iron clustered closely with the soil grown plants. The other iron regimes in the HS produced highly tillered, highly productive plants; these treatments did not appear to impact the morphology within the AL regional group, illustrating the continued import of AL origin in the less severe treatment regimes. In contrast, the high iron treatment tended to produce plants that were relatively short and unbranched, leading to some overlap in morphology between the different regional groups defined by DBSCAN.

CONCLUSIONS

Here, I illustrate that while it is possible to influence the morphology of *Setaria* using an excess iron treatment, morphology and the ionome are disconnected in the *Setaria* species complex when these plants are grown in the soil. Additionally, I suggest that the varieties of this crop that have been exposed to high levels of phosphate during their breeding have developed a constitutive phosphate deficiency response.

Figure 1-1

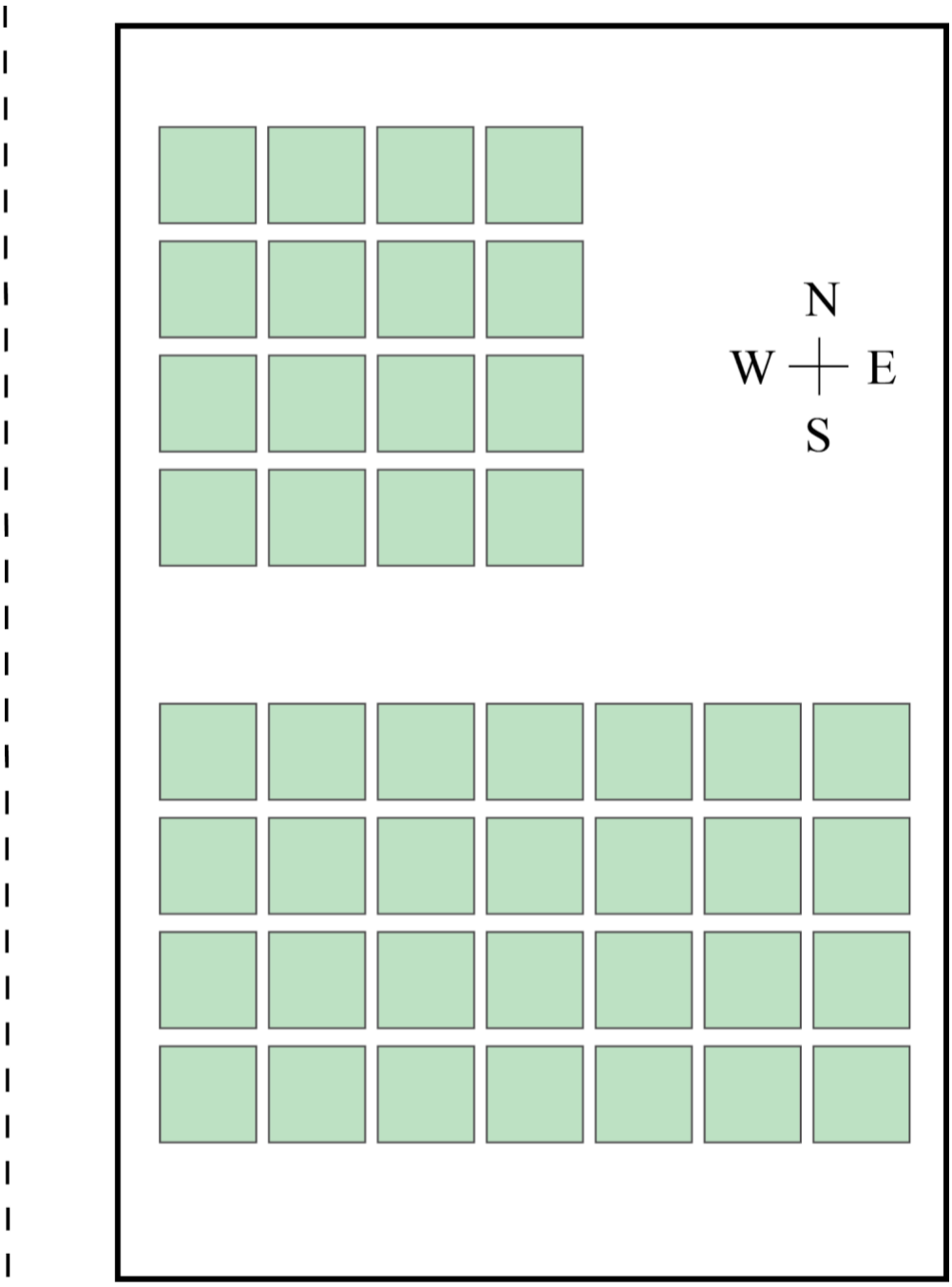


Figure 1-1. Arrangement of units in the hydroponic system

Plants were grown on a moveable bench approximately 1-3 meters from the west wall of the greenhouse, represented here by a dashed line. Boxes represent individual hydroponic units.

Figure 1-2

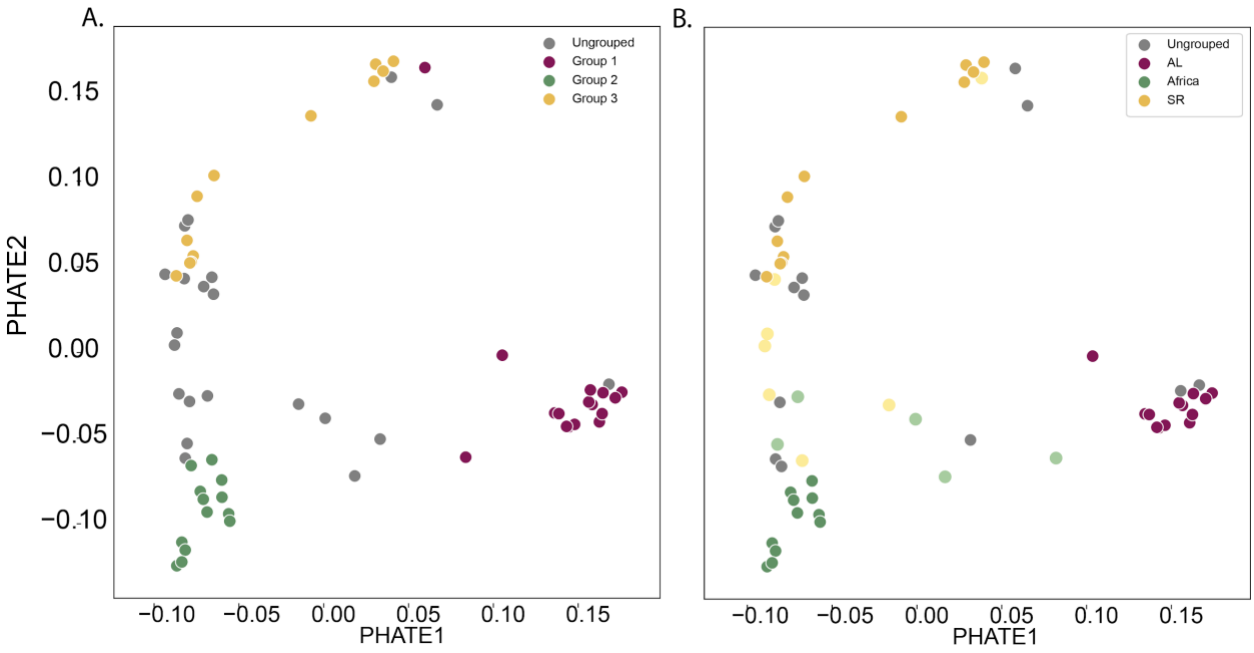


Figure 1-2. PHATE visualization of the morphologically defined groups assigned to soil plants

The morphological groupings produced by DBSCAN (A) correspond well to the groups designated by origins in Afghanistan/Lebanon (AL), Africa, and China/Russia (here termed Sino-Russian) (B). Lighter colors indicate the projected clustering of previously ungrouped data points based on accession.

Figure 1-3

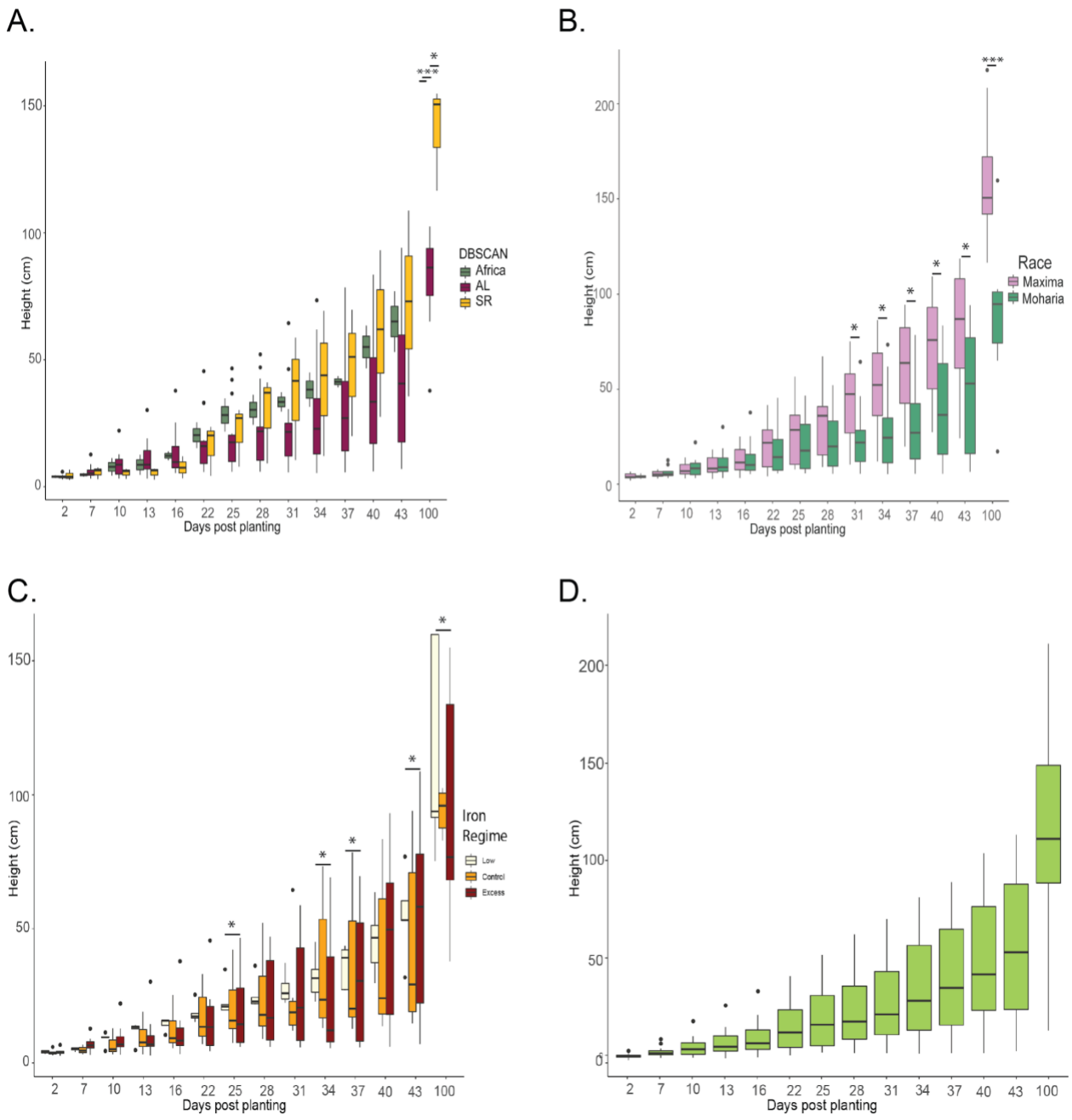


Figure 1-3. Time course of height in the HS

Figure 1-4

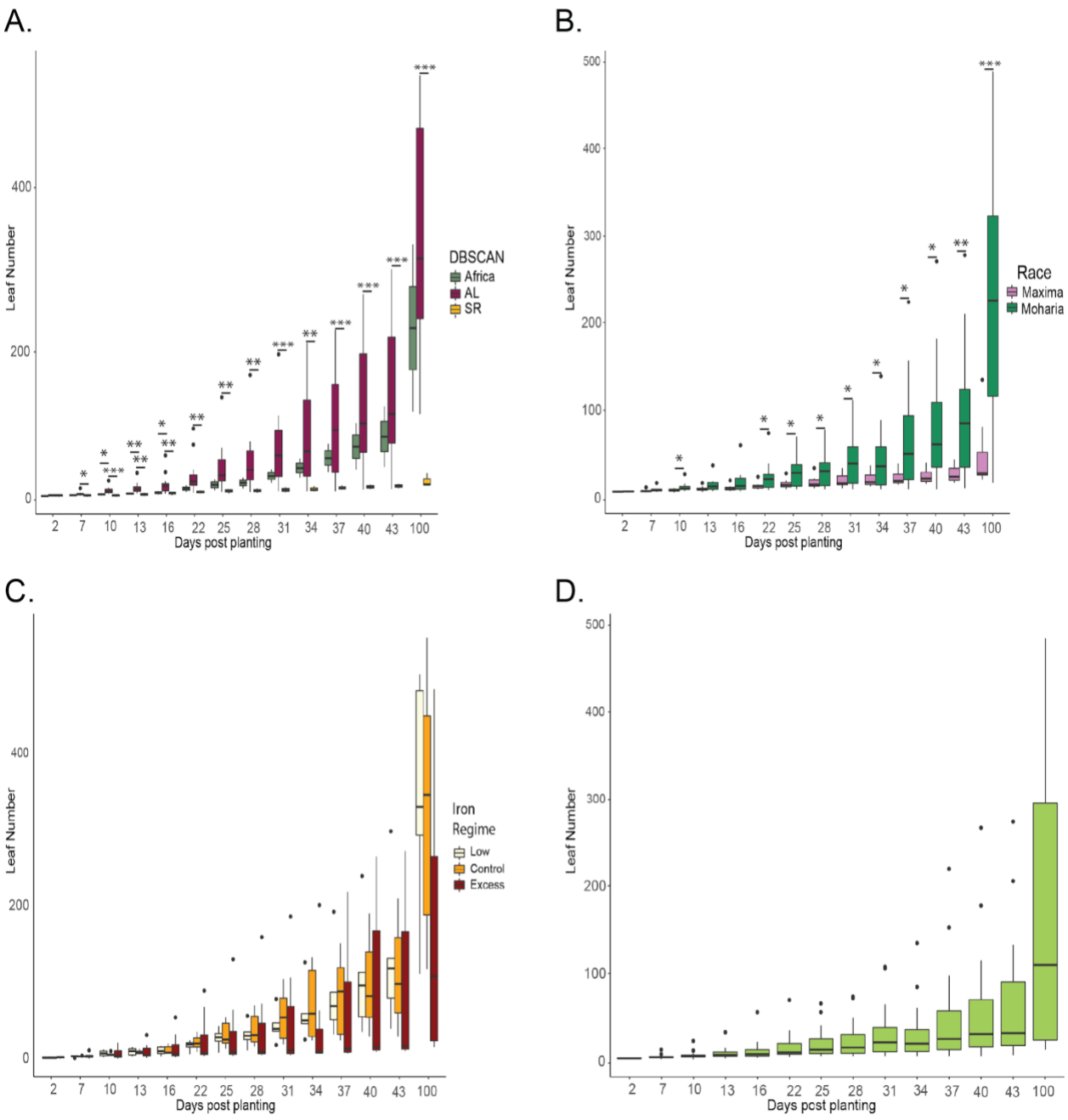


Figure 1-4. Time course of leaf number in the HS

Figure 1-5

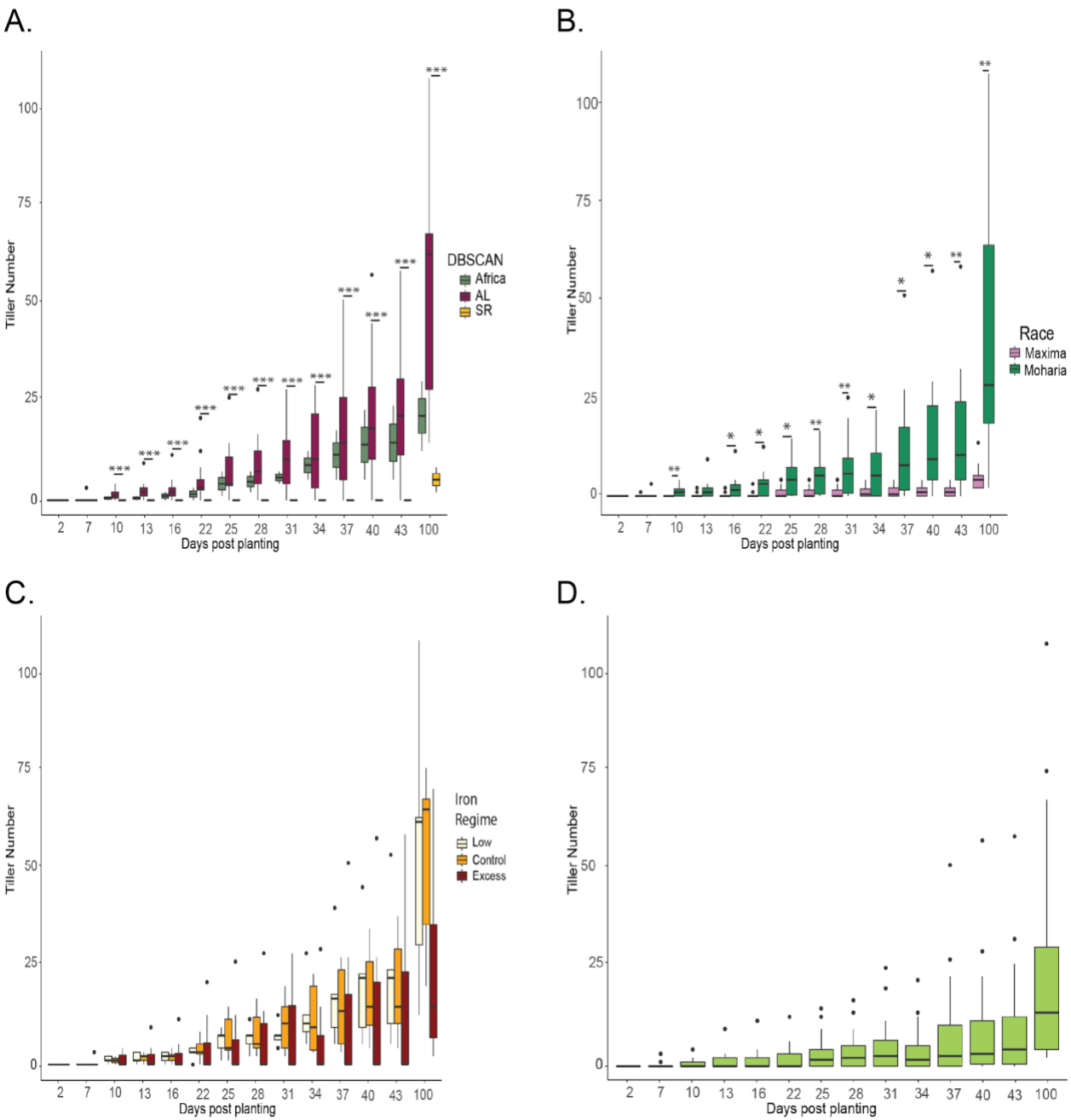


Figure 1-5. Time course of tiller number in the HS

Figure 1-6

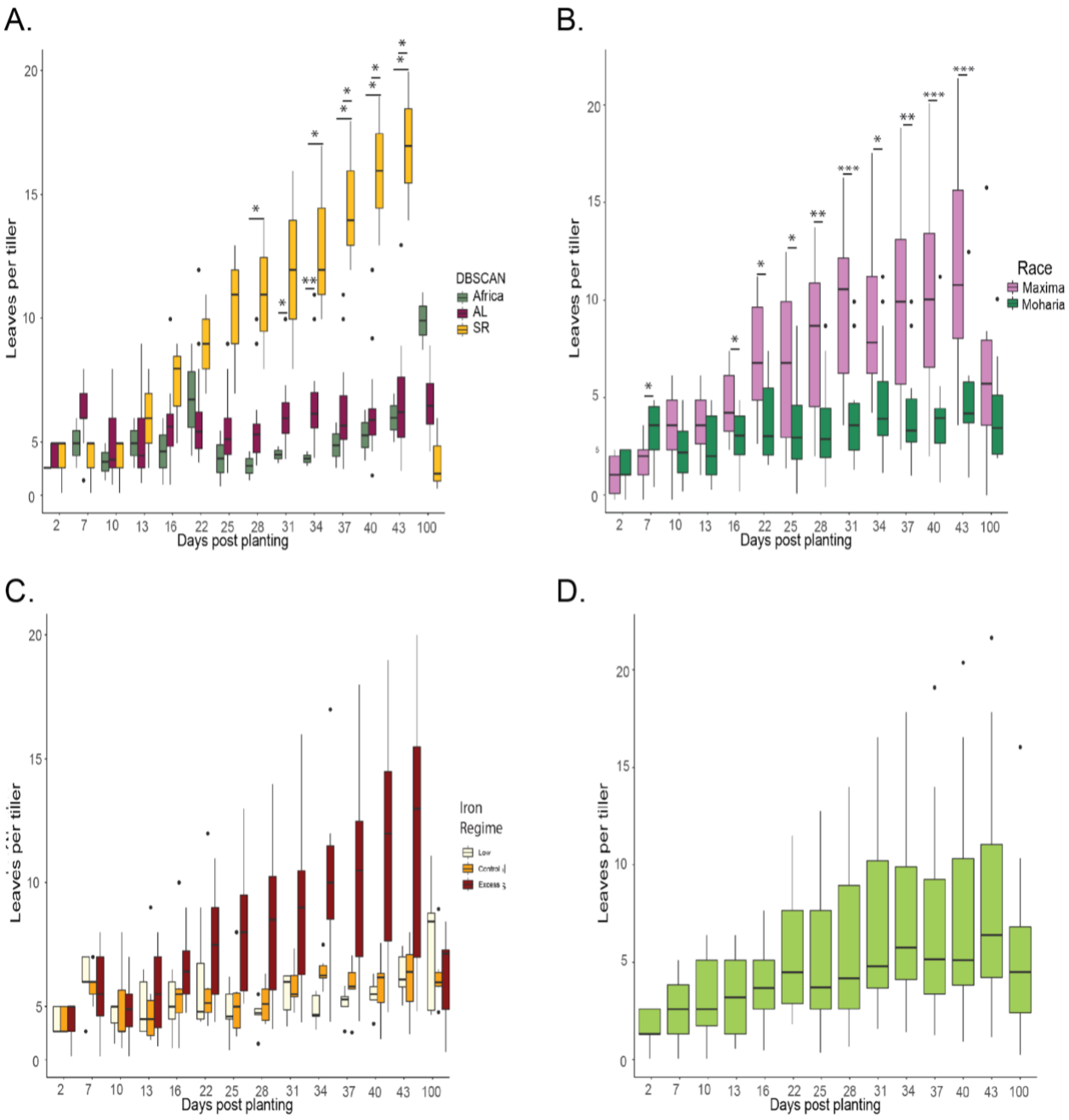


Figure 1-6. Time course of average number of leaves per tiller in the HS

Figure 1-7

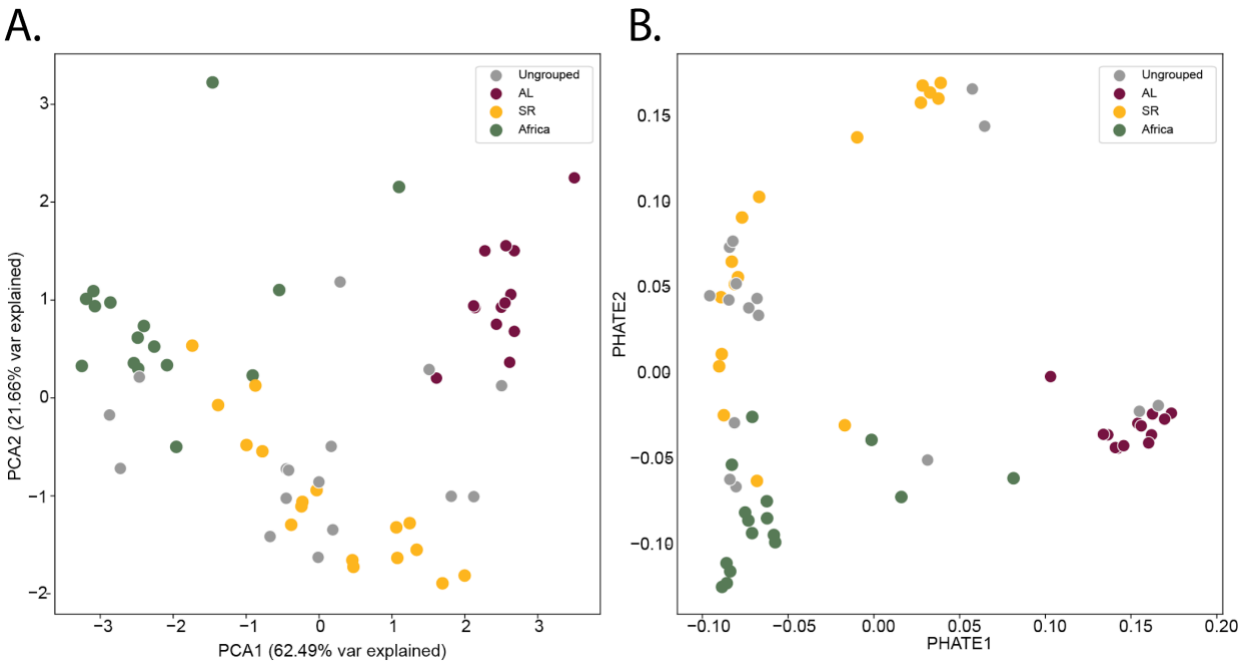


Figure 1-7. PCA and PHATE of morphological dataset for soil grown plants

Both methods clearly illustrate the result of DBSCAN morphological groups.

Figure 1-8

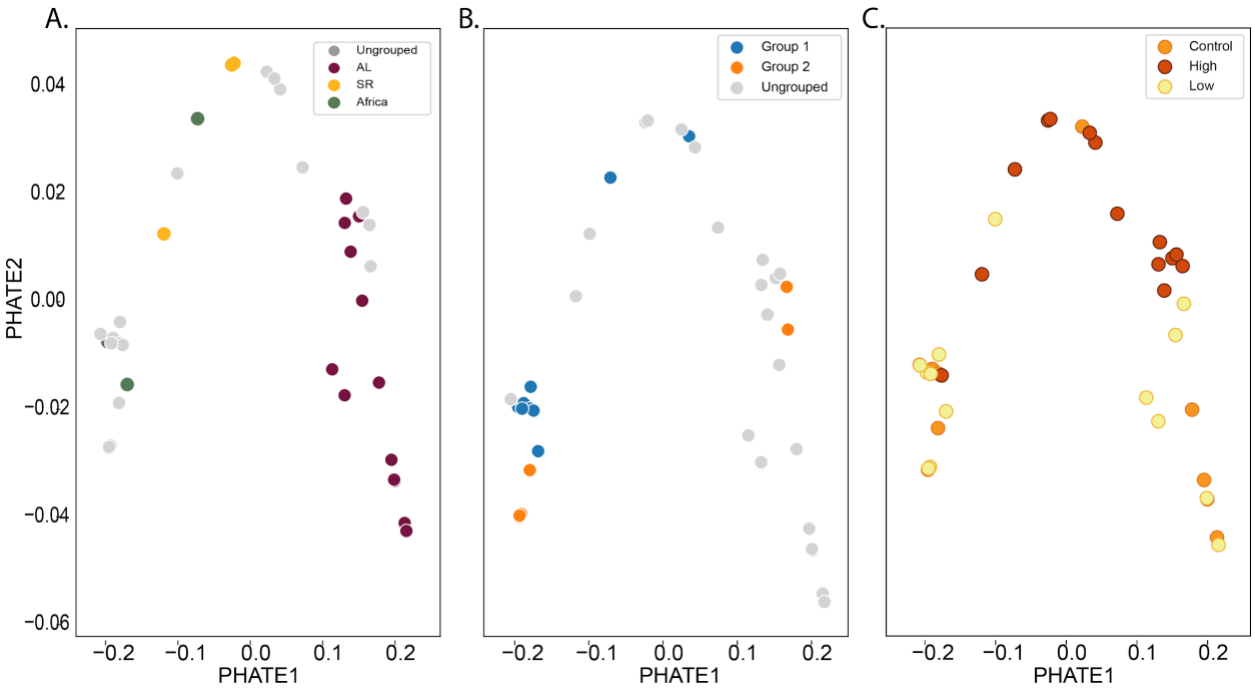


Figure 1-8. PHATE visualization of the morphology of hydroponically grown plants

HS plants were best grouped by treatment (C). Ionomically defined groupings were not clearly present (B), and of the clusters defined by the morphology of soil plants, only AL seemed to consistently cluster in the rightmost arm of the plot (A).

Figure 1-9

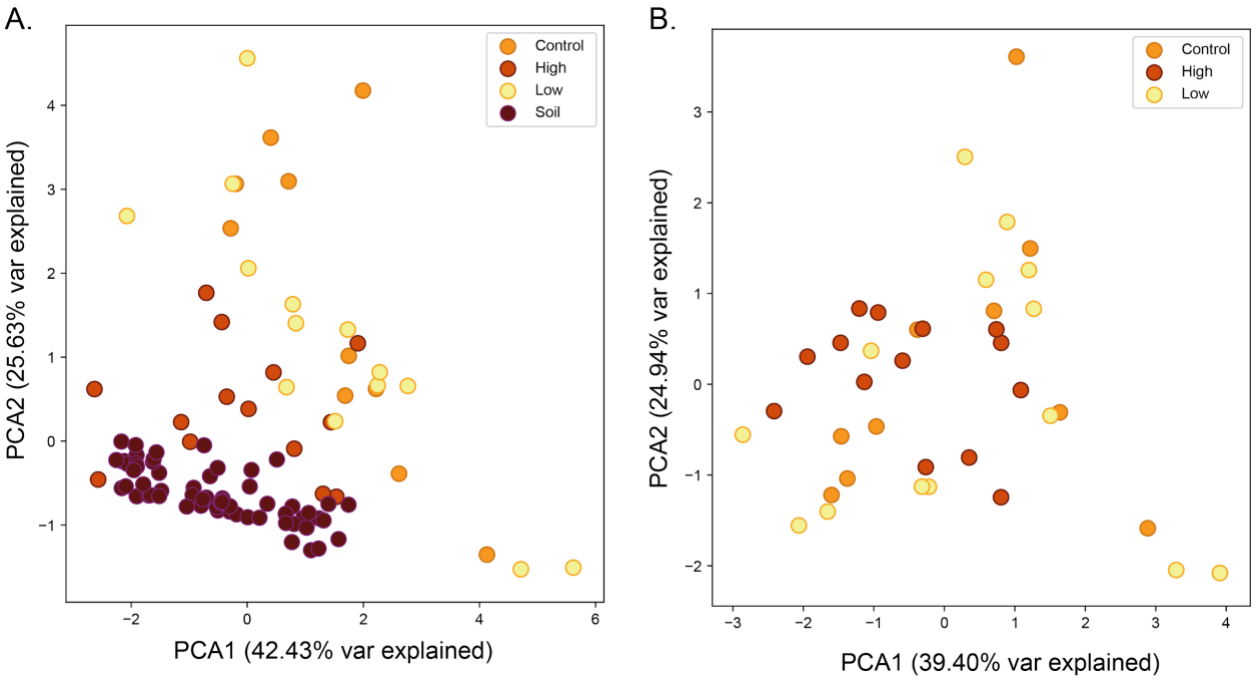


Figure 1-9. Plant morphology datasets cluster by treatment

HS plants in the high iron treatment tended to cluster near the origin (B). Plants in the other two HS conditions did not appear to cluster by treatment. When the HS and soil grown plants' morphology were compared, plants treated with excess iron tended to cluster with the soil grown plants (A).

Figure 1-10

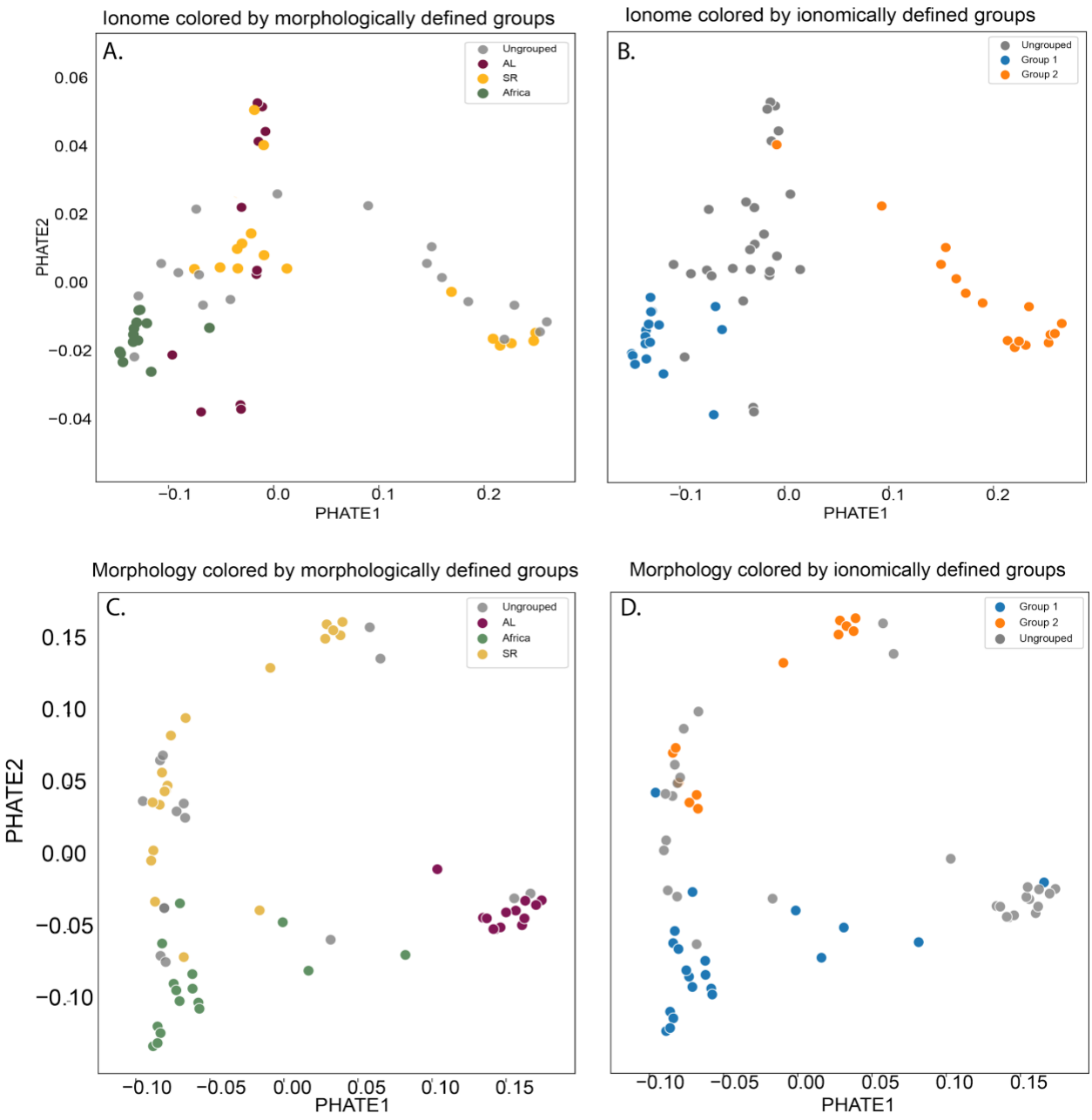


Figure 1-10. The DBSCAN algorithm identifies disparate clusters in the ionomic and morphological datasets

The ionomic clustering (B,D) does not follow the same pattern as the morphological clustering (A,C) with the exception of the African morphology group, which appears to correspond to the ionomic cluster Group 1.

Figure 1-11

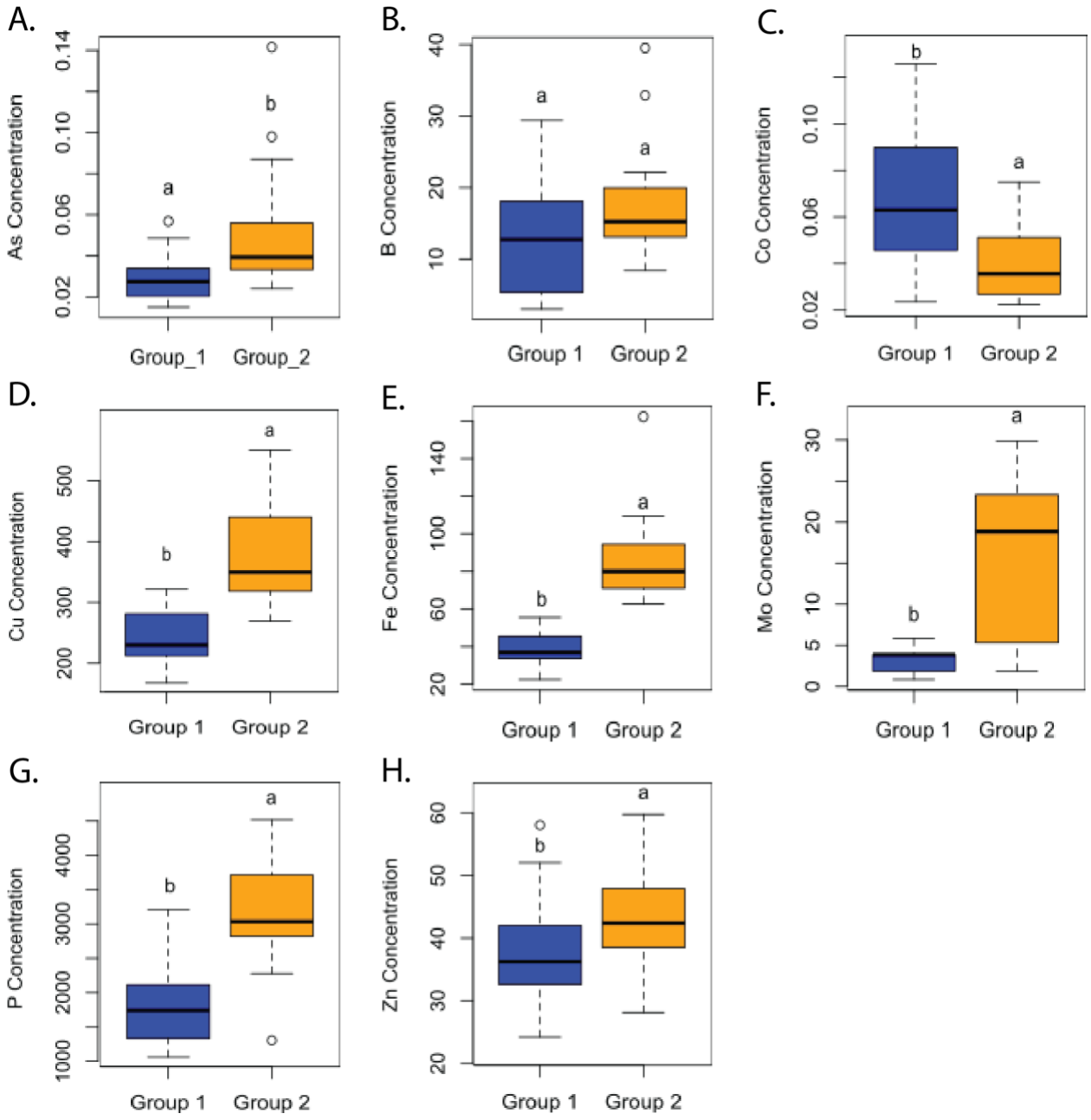


Figure 1-11. Comparison of the concentration of ions indicative of phosphate deficiency in soil grown plants

Group 1 and 2 were compared. All ions except boron were significantly different in concentration.

Figure 1-12

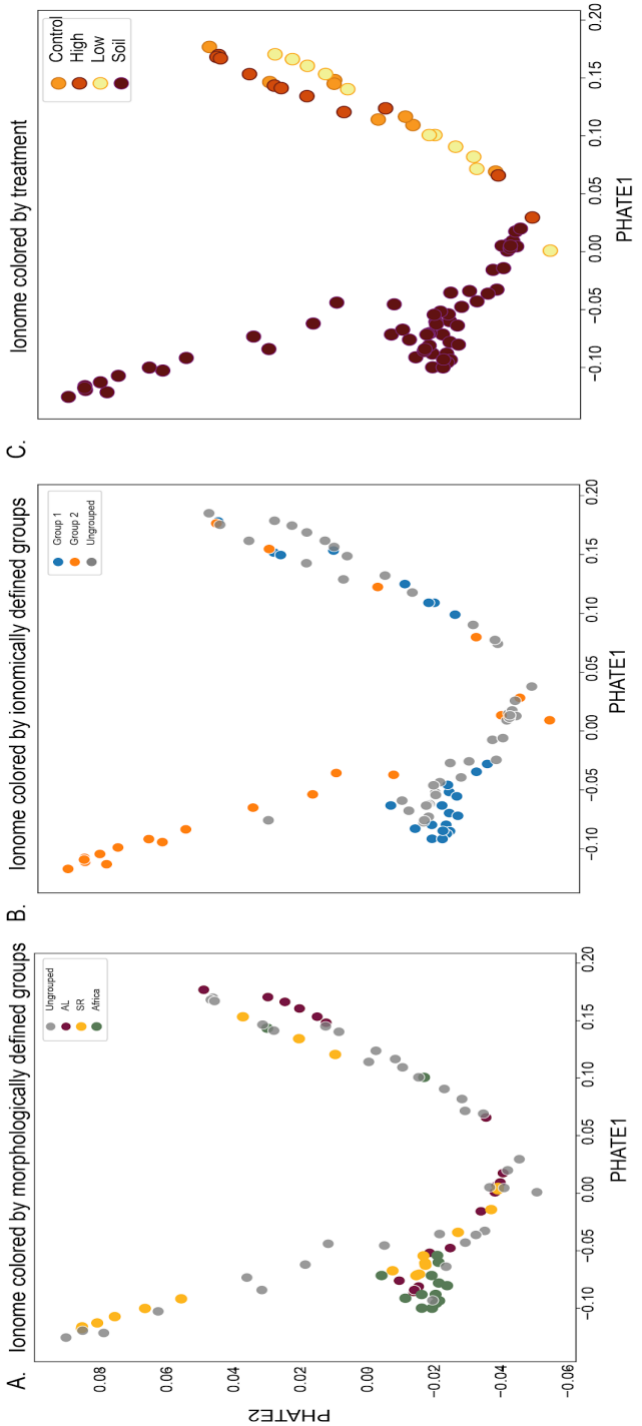


Figure 1-12. The ionome is perturbed by HS conditions

PHATE visualization illustrates that plant ionomes cluster by treatment regime (C), with soil grown plants forming one cluster and HS plants forming a second.

Table 1-1

Race	Tillering	Height	Origin	Inflorescence architecture	Days to flowering
Maxima	1-8 (av. 1.6), unbranched	45-100 cm	China, Mongolia, Japan, Korea, Georgia. Introduced to the US.	East Asian: 12-30 cm, pendulous. Well developed bristles. Central Asia: shorter, erect. Short bristles.	32-57 dpp
Moharia	5-52 (av. 8.6), heavily branched	25-100 cm	Europe, Russia, Afghanistan, Pakistan	Erect or nodding. Short branches. Bristles tend to stand out of the panicle.	32-50 dpp
Indica	1-25 (av. 3.4), some branching.	Variable	Appears to be a cross between *maxima* and *moharia*. Largely grown in India and Sri Lanka.	6-30 cm long. Erect or nodding (if larger). Well developed bristles.	43-70 dpp
Setaria viridis	Extensive branching.	Maximum 150 cm, typically 30 cm in Berkeley	China	<10 cm. Long bristles. Easily shattered at maturity.	-

Table 1-1. The racial morphology of *Setaria*

Setaria has been divided into as few as two and as many as four racial subgroups. The three most commonly included racial groups are enumerated here, as is the wild *S. viridis.*

Table 1-2

Type	Description
Entisols	Little, if any horizon development
Aridisols	Soils located in arid climates
Alfisols	Deciduous forest soils
Ultisols	Extensively weathered soils
Gelisols	Soils containing permafrost
Andisols	Soil formed in volcanic material
Inceptisols	Beginning of horizon development
Mollisols	Soft, grassland soils
Spodosols	Acidic, coniferous forest soils
Oxisols	Extremely weathered, tropical soils
Histosols	Soils formed in organic material
Vertisols	Shrinking and swelling clay soils

Table 1-2. USDA defined soil types and descriptions

Note that several soil types, including gelisols and inceptisols, are largely nonarable.

Table 1-3

Origin	Species	Accession Number	Other Identifiers	Probable soil type in region	Notes
Hungary, Tapioszele, Pest	*Setaria italica*	531447	'SZENTESI "D"'	Mollisol, Alfisol	Cultivar
Andhra Pradesh, India	*Setaria italica*	464157	NA	Fluvisol, Entisol	Locality: Andhra Pradesh State Latitude: 16 deg. 15 min. 50 sec. North (16.26396530), Longitude: 80 deg. 45 min. 32 sec. East (80.75906360)
Afghanistan	*Setaria italica*	207502	NA	Likely Entisol or Aridisol	
Zhang Nong	*Setaria italica*	642744	Zhang Nong No. 10	Entisol, Inceptisol	Breeding material
China	*Setaria faberi*	669940	'Ames 26106'	Unknown	Cultivar
USA, South Dakota	*Setaria italica*	649317	'Manta'	Mollisol, Entisol, Alfisol	Elite Cultivar
Kenya, possibly Nairobi	*Setaria sphacelata*	165718	NA	Andosol, Cambisol, Nitisol	
Argentina, Cordoba	*Setaria parviflora*	509031	NA	Entisol	Locality: Route 8, 132 km W of Venado Tuerto Latitude: -33 deg. -23 min. -30 sec. South (-33.39190000), Longitude: -63 deg. -18 min. -14 sec. West (-63.30410000)
Nangarhar, Afghanistan	*Setaria italica*	212626	NA	Inceptisol, Entisol	Locality: Market collection in Jalalabad Latitude: 34 deg. 22 min. 58 sec. North (34.38290000), Longitude: 70 deg. 32 min. 21 sec. East (70.53930000)
Afghanistan	*Setaria italica*	253495	NA	Likely Entisol or Aridisol	
Kenya	*Setaria italica*	464554	NA	Unknown	
USA, Colorado	*Setaria italica*	649318	'White Wonder'	Alfisol, Mollisol	Elite Cultivar; Observed in Minnesota: Late maturity. Good lodging resistance. Reported to be less drought resistant than other varieties, but not observed. Small white or yellow seed. Too late for good seed production some years.
Unknown	*Setaria viridis*	A10.1	'A10'	Unknown	Lab strain
Northern China (Sun et al 2019)	*Setaria italica*	None	'Yugu', 'Yugu No.1'	Unknown	Elite cultivar
Kenya	*Setaria italica*	464555	NA	Andosol, Cambisol, Nitisol	
Georgia, Ames	*Setaria italica*	32317	NA	Alfisol, Mollisol, Inceptisol	Locality: Village Magliki, Tskaltubo District, Imereti Province, Caucasus Region. Latitude: 41 deg. 37 min. 28 sec. North (41.62453333), Longitude: 42 deg. 43 min. 54 sec. East (42.73171667)
Bolivia, Tarija	*Setaria macrostachya*	509036	NA	Ultisol, Oxisol	Wild material; Locality: 44 km E of Boyuibe Habitat: Dry area Latitude: -20 deg. -28 min. -55 sec. South (-20.48200000), Longitude: -62 deg. -50 min. -56 sec. West (-62.84890000)
Spain	*Setaria italica*	283988	NA	Inceptisol	
Lebanon	*Setaria italica*	452244	NA	Leptosol, Cambisol	
South Africa, Transvaal	*Setaria italica*	209909	NA	Oxisol, Inceptisol, Entisol	
Poltavskaya, Russia	*Setaria italica*	315090	NA	Mollisol	Cultivar

Table 1-3. ARS GRN accessions selected for hydroponic assay

Plants were selected based on diversity in soil type at origin, region of origin, and breeding status.

Table 1-4

Feature	Feature Importance
Average panicle weight	23.90
Average panicle length	20.45
Height	16.25
Total panicle weight	15.93
Tiller number	14.33
Leaf number per tiller	9.13

Table 1-4. Morphological feature importance for the random forest regression assigning DBSCAN groups

Table 1-5

Ion	Feature Importance
Fe	16.76
Mo	10.80
Na	8.46
Sr	6.64
S	5.79
Mg	5.76
K	5.39
As	5.12
Cu	4.59
Ca	4.43
Mn	3.47
Se	3.46
P	3.42
Co	3.31
Rb	2.51
Zn	2.48
Al	2.27
Cd	1.91
B	1.89
Ni	1.57

Table 1-5. Ionomic feature importance for the random forest regression assigning DBSCAN groups

Section Two. Identifying the genomic regions associated with the ionome in foxtail millet

PREFACE

As a result of the analyses conducted in Chapter 1, it became apparent that there was sufficient ionomic variation in the *Setaria* germplasm to carry out studies dedicated to the mapping of the genetic loci associated with ionic homeostasis. The plant ionome is a highly multidimensional trait that involves the complex choreography of up to 20 unique ions. Additionally, the loci responsible for the interaction *between* elements must be identified. This intercorrelation is easily captured by means of a data dimensionality reduction method such as principal components analysis, the results of which are used as an additional phenotype for mapping. Any ionomic mapping study is therefore susceptible to issues related to multiple testing. That is to say, it is likely that a large proportion of the genome will appear to be associated with a trait of interest if a sufficiently large number of traits are assessed. The question therefore arises: How can this particular multiple testing issue be addressed? In Chapter 2 I seek a statistically rigorous method to answer this question. The resulting statistical tool, the SPQV, is utilized in Chapter 3, wherein a QTL mapping study is finally employed.

Chapter 2. The Scanning Probabilistic QTL Validator

INTRODUCTION

Quantitative trait mapping identifies regions of a genome, known as quantitative trait loci (QTL), that are associated with a phenotype of interest (Miles and Wayne 2008; Tanksley 1993). This mapping process relies on data collection and extensive curation: the collection of data, the removal of both global and conditional outliers, and model selection are all necessary features of the protocol (Miles and Wayne 2008; Liu 2017). All of these steps rely on the researcher's best judgment, and are therefore subject to bias (Kleyman *et al.* 2017; Collett and Lewis 1976; Broman 2001). QTL mapping requires significant effort and resources, and often serves as a starting point for even larger projects in plant breeding or fine-mapping alleles that contribute to phenotypes (Yin *et al.* 2003; Collard *et al.* 2005). It is therefore important to be confident in the quality of QTL mapping results and analyses.

Currently, the quality of a QTL mapping experiment is assessed by cross validation (CV) and functional Gene Ontology (GO) term enrichment analysis (Würschum and Kraft 2014; Yon Rhee *et al.* 2008; Salih and Adelson 2009). In CV, QTL detection is performed using a subset of the lines involved in the original experiment to assess the robustness of the results; though this method is a valuable tool, there remain concerns about the influence of the population structure on the effectiveness of CV (Würschum and Kraft 2014). GO term enrichment analysis is another method by which the location of QTL can be validated. This method is designed to determine if the annotations of the genes in the identified regions are more frequently associated with the phenotype of interest than are the annotations of the genome as a whole (Yon Rhee *et al.* 2008). This approach relies on both accurate GO term annotation and the researcher's understanding of the complex processes that may contribute to the phenotype.

A third method for the assessment of QTL leverages our knowledge of genes that have previously been associated with the trait of interest (hereafter referred to as 'known genes') either in the organism of interest or in its close relatives (Doust 2004; Odonkor 2018). These genes can act as a sort of 'sanity check', as known genes of large effect should be found within the identified QTL. Single genes can underlie QTL, but multiple, related genes may also act in concert to control the placement of a QTL (Studer and Doebly 2011). Because the confidence intervals for QTL can be placed some distance away from the causative locus, it becomes necessary to determine if the number of known genes that are found within the confidence intervals of a QTL can be considered statistically significant.

One strategy that can allow us to understand the likelihood of identifying a particular number of known genes, given the null hypothesis of random placement of QTL, is resampling with replacement (RWR; Efron and Tibshurani 1986). RWR analysis for this purpose begins with the random selection of regions of the genome that are equal in size to the identified QTL. In each repetition, one records the number of known genes that are found within the selected region.

After a large number of repetitions ($n \geq 1000$), confidence intervals for the distribution of values produced by RWR can be calculated. These confidence intervals can be compared to the observed number of known genes found; if the observed number exceeds the calculated confidence limit, the QTL can be considered to have found a significant number of known genes.

QTL RWR is sensitive to many different factors. Among these are the restrictions that are imposed on QTL placement, the treatment of closely linked genes, and the method chosen to handle the constraints imposed by the physical properties of the genome. Additionally, RWR assumes that the distribution of known genes reflects the true distribution of both known and unknown genes associated with the trait of interest. The present work aims to identify an optimal strategy for RWR analysis of QTL mapping studies. We ultimately propose a new method, Scanning Probabilistic QTL Validation (SPQV), that is designed to overcome the pitfalls associated with the simplest instance of RWR, specifically in its assumptions surrounding the gene distribution. We discuss the assumptions made in SPQV versus RWR in the context of the reference genome of *Setaria italica*. Both methods are used to analyze the results of a simulated QTL mapping experiment. Finally, the SPQV is used to analyze the results of a previously published QTL mapping study.

Resampling with replacement: An apparently simple method

RWR validation of a given QTL involves resampling the genome for randomly positioned regions of length *L* a large number of times, where *L* is the distance between the left and right confidence intervals of the QTL of interest. In the current work, all distances are measured in terms of base pairs. The statistic for each run is *K*, the number of known genes in the chosen region. In the simplest instance of RWR, the region of length *L* has no restrictions on location, other than that it must be placed fully on a chromosome. The random selection of the region of length *L* starts with the selection of an origin base pair, *O*. The QTL then extends outward from this point to cover the chromosomal interval [*O*, *O+L*]. Some complications arise at the tail ends of chromosomes, for example: if a chromosome is length *C*, then the region contained in [*C-L*,*C*] can never contain *O* and will therefore have a reduced likelihood of being included in the assessment (*Figure 2-1*).

This ‘underrepresentation’ of the tail ends of chromosomes is non-negligible due to the distribution of genes on individual chromosomes. Generally, the number of genes increases with increasing distance from the centromere, and drops precipitously at the relatively short telomere (*Figure 2-2*; Paape *et al.* 2012). There are several methods that can be used to handle the issue of underrepresentation, the first of which can be termed the ‘bounceback method’ (*Figure 2-1A*). In the bounceback method, if *O* is selected such that the region of length *L* extends past the end of the chromosome, *O* is placed on the last base pair of the relevant chromosome, with the selected region extending to the point *C-L*.

The bounceback method does rectify the underrepresentation of [*C-L, C*], but results in *overrepresentation* of the tail ends of chromosomes.

A second method for addressing the issues associated with the assessment of the tail ends of chromosomes involves the directionality of QTL extension. In the case of unidirectional QTL extension, the origin *O* of the region of length *L* is chosen, and the selection for the RWR sample is then considered to be [*O, O+L*]. Unidirectional QTL extension's most prominent fault is in its treatment of the chromosome as a string of text, rather than as a physical object. Mapping in this manner biases gene discovery, particularly for longer QTL, as genes towards the 'tail' end of the chromosome are less likely to be found within *L*. Bidirectional QTL extension (in which the selection for the RWR sample is defined as [*O, O*$\pm$*L*], with the operation selected at random when possible) reduces the underrepresentation of the tail end of the chromosome by allowing *O* to fall into the region [*C-L, C*]. Unfortunately, a slightly less serious issue remains: on a chromosome of length *C*, the origin O continues to have a somewhat reduced likelihood of falling in the regions [1, 1+*L*] and [*C-L, C*] as compared to the region [1+*L*, *C-L*].

Selecting *O* at random is at the heart of RWR, but some restrictions on the location of *O* are necessary to reflect the actual process of QTL mapping. In general, QTL mapping starts with the identification of the marker that appears to correlate strongly with the trait of interest. The outer boundaries of the QTL are then defined as the region in which the causative locus lies with 95% certainty; these boundaries can then be extended to the closest markers used in the mapping process (Broman *et al.* 2003; Arends *et al.* 2016). In order to best reflect this process, the placement of *O* should be limited to the markers used in the original mapping experiment. This is particularly important because markers, like genes, are not evenly distributed through the genome, with decreasing density near centromeres and telomeres. Bidirectional QTL extension becomes more important in this context, as the genes located towards the boundary of an interval between markers have either disproportionately high or low likelihoods of being identified, depending on their placement (*Figure 2-1C, D*).

Because QTL mapping is performed on organisms with more than one chromosome, and because RWR relies upon the distribution of known genes, it is important to recognize the uneven dispersion of the known gene list (*Figure 2-2*). Duplication events often result in genes of similar function in tandem array (Reams and Neidle, 2004; Fan *et al.* 2008), and genes in functional groups appear to cluster on a larger scale as well (Chuang and Li 2004; Salih and Adelson 2009). If the QTL found in the original mapping experiment were required to span the entire distance between markers, any genes that are not separated by markers will function as one genetic unit - that is to say, it is impossible to identify one without identifying the others *(Figure 2-3).*

Due to linkage events of this sort, it is possible to produce a confidence limit via RWR that can only be exceeded by the identification of one specific, disproportionately gene dense locus, even though a much more relaxed interval would have been produced with the use of a single extra marker that split the tandem array. Because of this, known genes that are not separated by markers should be treated as a single locus during RWR. Linkage should also be taken into account when comparing the results of an experiment to the confidence limit produced by RWR: if multiple genes contained between the same two markers were identified in the mapping experiment, they must be treated as a single locus.

Once RWR is performed, confidence intervals are still complicated to calculate, because the distribution of genes found is not smooth and is not normally distributed (Figure 2-3). Because standard confidence intervals rely on the assumption of normality, they cannot be used with the distributions typically produced via RWR. The use of Bias Corrected and accelerated confidence intervals is recommended (Efron 1987; Efron and Narasimhan 2018), which requires fairly sophisticated math to implement.

RWR has the potential to be a powerful method for the assessment of QTL mapping, but its execution requires thoughtfulness on the part of the individual researcher. Different execution strategies also make it difficult to compare validation results between papers. Additionally, there remain some problems with the RWR method, particularly its reliance on the known gene distribution, which is unlikely to reflect the true distribution of genes associated with the trait. It is also apparent that even with the bidirectional extension of QTL there remains a tendency to overrepresent the centers of chromosomes. Given the central importance of gene distribution to RWR, addressing this particular failing will produce a method that is much more effective at identifying QTL of interest, and will therefore improve the rate at which breeding and fine mapping can be accomplished.

MATERIALS AND METHODS

Using the Scanning Probabilistic QTL Validator to validate mapping experiments

In light of the problems discussed above, we propose a new method for the analysis of QTL mapping experiments. Termed the Scanning Probabilistic QTL Validator (SPQV), this method uses probabilistic methods to determine the likelihood of finding a particular number of genes, *K*, given the null hypothesis of random placement of QTL. In order to answer this question, we calculate the Expected Gene Number (EGN) that accounts for the probability of randomly finding previously-identified genes.

This EGN and its confidence interval is completely determined by five factors:

1. The number and lengths of QTL that were identified for a given trait
2. The *number* of previously known genes related to the trait of interest
3. The locations of markers used in the mapping experiment
4. The locations of all genes in the genome
5. The physical architecture of the genome (i.e., the locations of chromosome ends)

What is the probability of identifying a gene? The probability of identifying a particular gene depends upon the likelihood of placing the QTL of interest on any of the markers in range of the locus. The range is defined as $[\ell - L\ ,\ \ell + L]$, where ℓ is the gene's locus (modeled here as a point value corresponding to the center of the gene's coding region) and L is the length of the QTL.

Addressing the distribution of markers in the genome: The probability associated with the selection of a particular marker, $P(m_i)$, as the origin for a particular QTL is calculated by assessing the number of locations in the genome on which the QTL can be placed. $P(m_i)$ is determined by the number of markers that can be used as the origin O of the QTL of length L, a number which excludes all markers that would result in the QTL extending beyond the end of the chromosome. This origin-based model of QTL mapping is an approximation for the true process of QTL placement, wherein QTL are roughly centered on a marker and are terminated at a marker on either end which represent the 95% confidence intervals for that QTL. In the case where the original QTL were not terminated at markers, it is preferable to use a model in which the QTL is centered on the chosen marker. Models requiring that the QTL must both start *and* end on a marker are not feasible, because the distance between markers is not uniform; with this constraint, a QTL of a given length might have only one possible genomic location.

Markers are used in a direction-independent manner to avoid underrepresentation of the ends of the chromosomes. Under the null hypothesis, the probability of using any marker is $\{0,1,2\}/M$, which can take on any value between 0 and 1, inclusive. Here, the numerator depends on whether the marker can serve as O with unidirectional QTL extension, bidirectional QTL extension, or with neither; M refers to the total number of markers from which the QTL can extend to the left plus the total number of markers from which the QTL can extend to the right (*Figure 2-4A*).

Although the denominator could be adjusted to take into account the fact that a single QTL can only be mapped to one chromosome, this is unnecessary, because a QTL is not equally likely to be mapped to every chromosome. To understand why, consider the most obvious strategy to account for differences in chromosome length and the number of usable markers: a weighting scheme. We would divide the number of usable markers on each chromosome (M_C) by the total number of usable markers in the genome (M).

If this weighting is then used to adjust the probability $P(m_j)$ of using any marker on that chromosome, which is already proportional to $1/M_C$, the chromosome-specific marker counts cancel out and leave only the whole-genome marker count.

Designing simulated known gene distributions

To best represent the topology of the genome, the SPQV simulates genetic loci by selecting genes at random from the whole genome gene distribution to represent the genetic basis of the trait of interest. The use of the whole genome gene distribution as a source accounts for the topology of the genome, including the decrease of gene density at the centromere and telomeres (*Figure 2-5*).

This strategy assumes that the true distribution of genes associated with a particular trait is approximately the same as the distribution of genes on a whole, rather than the assumption used in the simplest instance of RWR: that the distribution of known, previously associated genes reflects the true distribution of genes associated with that trait. We argue that this novel assumption is more likely to be accurate because trait-related genes can easily be discovered in a spatially biased manner (Koren *et al.* 2007): tandem arrays promote clustered discovery, some transposons involved in transposon-mediated mutagenesis preferentially target certain sequences (Kawakami *et al.* 2017), and genes in regions close to the centromere tend to be difficult to identify through methods (like QTL mapping) that rely on recombination (Noor *et al.* 2001). Additionally, the genome is interconnected; many traits rely on the interaction between multiple, seemingly disparate biological processes. Use of a random distribution for simulating genes with the SPQV is also possible, but fails to capture the genomic topography. Because genes within functional groups are not randomly arranged, duplication events and gene clusters in the original set of known genes are taken into account by considering genes without a marker between them as one genetic unit.

Calculating EGN: The markers within range of each locus in the simulated distribution are found and their associated probabilities summed in order to determine the probability of randomly detecting a gene at that locus, $P(g_i)$. $P(g_i)$ is then summed for all loci in the simulation to produce a single value. This value corresponds to the mean expected gene number for that specific QTL (Eq. 1). Given a sufficiently large number of simulations, the standard error of the mean (SEM) can be calculated. The 95% confidence interval can then be calculated for the distribution of mean EGN using the formula $\bar{\mu} \pm z_{a/2}SEM$. This value is then compared to *K*, and a determination can be made about the significance of the number of genes identified in the QTL at question.

We have produced an open-source R package "SPQV" that contains the necessary functions for the use of this method. The output of the *SPQValidate()* function includes both the lower and upper bounds for the 95% confidence intervals for EGN in each QTL.

To account for the presence of multiple QTL, the distributions of mean EGN for all QTL within a set[1] are combined. Because these distributions are normally distributed, the combined set has a mean equal to the averaged means of each distribution. To get the confidence interval for the combined set, CI_{tot}, we calculate a standard deviation by summing the variances of each individual distribution: $\sigma_{tot} = \sqrt{{\sigma_1}^2 + {\sigma_2}^2 + \ldots}$. Note that this process treats each QTL as if the available number of genes was the same when they were identified in the original experiment. In reality, each successive QTL likely has a lower likelihood of identifying known genes, as the earlier QTL mapped tend to be placed on highly significant regions. The multiple testing correction for the SPQV is therefore overly conservative for later QTL.

RESULTS

Simulated comparison of the SPQV and RWR

To assess the efficacy of the SPQV as compared to RWR, 200 QTL were simulated with lengths spaced logarithmically between 5,000 to 43,744,866 base pairs. This range was selected to span from the smallest QTL found in a typical experiment to the size of the largest full chromosome in *Setaria italica*. The largest QTL was dropped from the analysis, however, since it only fits in one place on the genome, leaving 199 QTL. A simulated "known gene" dataset of 500 genes was randomly selected from the *S. italica* genome, which was reduced to 314 genes after those not separated by a marker were condensed to one representative locus (*Figure 2-6*). A relatively high number of "known genes" was used in order to get more fine-grained comparisons between the methods. The experiment was also repeated for a real known gene list containing 71 genes associated with the plant ionome *(Supp. Figure 2-1,2).*

The simulated data was then used to calculate 95% confidence intervals using both SPQV and 8 different versions of RWR. The RWR experiments varied in the permitted QTL origins, directionality of QTL extension, and the use of bounceback (*Figure 2-7*). The confidence intervals resulting from RWR were Bias Corrected and accelerated confidence intervals (BCa). BCa confidence intervals are generally considered superior to standard and quantile intervals in cases where the distribution of RWR estimated values does not conform to the normal distribution (Efron 1987; Efron and Narasimhan 2018). In these experiments, the distribution of identified genes was often zero-inflated and was occasionally completely binary (*Figure 2-3*).

1 In this case, a "set" refers to a group of QTL that were calculated at the same time. For example, within one experiment with multiple treatments, where the phenotypes of individuals within each treatment are used to calculate QTL, at least two sets of QTL exist, with one associated with each treatment.

The SPQV values are clearly most similar to those produced by the RWR experiment that was closest to biological reality: marker-only QTL origins, bidirectional mapping, and no bounceback *(method 8, Figure 2-7).* This makes sense, as the SPQV method is designed as a smoothed version of an experiment with these characteristics.

Restriction of QTL origin to the markers that were used in mapping leads to an increase in EGN for RWR *(Figure 2-7)*. This effect occurs for all QTL lengths. It is likely that the increase of identified genes in the context of restricted QTL placement is attributable to the physical structure of chromosomes: the markers selected for QTL mapping have a similar distribution to the genome wide distribution of genes *(Figure 2-5)*, and are therefore relatively sparse in gene-poor regions such as the centromere.

Similarly, the use of bounceback leads to an increase in RWR identified genes for all lengths of QTL, though this increase is particularly noticeable for some of the larger QTL (1.6×10^7 - 2.9×10^7 bp). It is likely that the relatively large number of genes situated close to the ends of chromosomes is the main contributor to the impact of bounceback on identified gene number. The effect of bounceback appears to diminish with bidirectional QTL extension. It is possible that this reduction is due to a smoothing of the distribution, as the occurrence of bounceback is effectively split in half over the two separate tails of the chromosome. The presence of long and short arms on chromosomes, and the corresponding lopsidedness of the gene distribution, might also contribute to this phenomenon *(Figure 2-5)*. The use of bidirectional mapping appears to result in fewer genes identified by RWR, though this effect is relatively minor when compared to the effects of origin restriction and bounceback.

In spite of the prominence of dark colors on the left side of the heatmap, the confidence intervals identified for small and medium QTL by the SPQV and by RWR are fairly similar regardless of RWR method *(Figure 2-7; Supp. Figure 2-3)*. The CIs in this range were consistently far below 1 regardless of method. In all, the SPQV 95% confidence limit for small QTL tends to be slightly smaller than the one produced by the RWR method that takes the same biological realities into account (method 8, *Figure 2-7*). However, this makes little difference in practice, because they are both less than 1: because observed gene counts are integers, EGNs from either method will be rounded up *(Supp. Figure 2-3)*. In other words, if an SPQV confidence limit is defined as 1.2, the QTL of interest must have an observed gene content of 2 or more genes to be considered significant during general use.

For larger QTL, SPQV values tend to outsize those produced by RWR. These large QTL approach the size of a full chromosome, and can indeed be larger than several chromosomes within the *S. italica* genome. It is the authors' opinion that this is not an overestimation for the true distribution of genes associated with the trait of interest, as the true distribution likely has more than the known number of genes.

Because of this, significance is unlikely for very large QTL, except for in the case of a true distribution of genes that is extremely uneven at the chromosomal level.

Domestication traits in maize: a case study

A reduction in tiller number is a classical domestication trait in maize (Doebley 1992). Modern maize lines have been bred to grow as single stalked plants to facilitate high-density planting, while the maize progenitor, teosinte, is highly tillered. The genetic network associated with tiller suppression is controlled by the *teosinte branched1* (*tb1*) gene that also controls several other aspects of maize morphology (Dong et al., 2019), including inflorescence and floral architecture. Several mapping populations made from crosses between the W22 maize inbred line and teosinte were recently described and used to map several domestication traits (Chen *et al.* 2019, Shannon 2012). As expected, several domestication traits associate tightly with *tb1* pathway. Here, we use the QTL reported by Chen *et al.* 2019 to illustrate the utility of the SPQV. Only the QTL with the same effect direction in both maize/teosinte mapping populations were assessed.

Seven genes (Table 1) closely associated with the *tb1* pathway in maize were located in the Zm-W22 NRGene 2.0 assembly (Springer *et al.* 2018, Dong et al., 2019, Chen et al., 2019) and analyzed using SPQV. Notably, these genes were selected based on their strong, known associations with the branching pathway in maize. Since only high-confidence genes can be used accurately with our method, if any gene were not truly associated with the trait of interest, its presence will render the SPQV more stringent than necessary. The QTL identified for the traits BARE (barren ear base), EB (ear branch number), GLUM (glume score), KRN (kernel row number), STAM (staminate spikelet) and TILN (tiller number) have previously been connected with the *tb1* pathway in maize. These QTL were therefore assessed in relation to the seven genes in Table 1. The markers found in the W22 x TIL01 RIL subpopulation were used to determine the base pairs associated with the CIs of these QTL. Where the end points of the QTL did not have an exact match to a marker, the next closest marker was used so as to mimic ‘extension’ style mapping. The results of this analysis are reported in Table 2-2. Four of the assayed QTL identified a gene from the *tb1* pathway, corresponding to four out of six of the represented traits.

Runtime analysis: To estimate the expected runtime of the SPQV, empirical experiments were performed with three different values for each of various parameters: number of repetitions, length of QTL list, length of known gene list, length of marker list, and length of QTLs (*Figure 2-8*). When one parameter was varied, the value for each of the other four parameters was set to the median value: 100 bootstrap repetitions, 3 identical QTLs of length 10 million bp, 15 known genes, and 3,000 markers. For most parameter settings, the analysis took less than a second, and no analysis took more than 10 seconds. Gene and marker lists were randomly sampled from the W22 maize genome for every run (Springer *et al.* 2018).

Each experiment was run 100 times using the microbenchmark library (Mersmann *et al.* 2019). The experiments were run with R version 3.6.2 on a 2019 MacBook Pro with a 2.3 GHz Intel Core i9 processor. Full details and code are available in the figures section of our GitHub repository.

DISCUSSION

Methods for usage: If the various adjustments described in this paper are applied to the RWR-based assessment of QTL mapping experiments, a more apt confidence limit for the expected number of genes will be identified. These adjustments do not account, however, for all of the issues associated with the application of RWR to this particular variety of question. RWR not only continues to rely on the distribution of known genes, but also results in gene-count distributions that nearly always fail to meet the requirement for smoothness (*Figure 2-3*; Davison and Hinkley 1997). These distributions, in other words, have a tendency to change abruptly, and are frequently binary in the case of small and very large QTL. The 'unsmoothness' of any given distribution will be exaggerated by small QTL size and short lists of known genes; a known gene list with fewer than one gene per chromosome, for example, would produce a binary distribution even for large QTL. Additionally, RWR continues to exhibit a reduced likelihood of the QTL falling in the regions [1,1+*L*] and [*C*-*L*, *C*] even with the adjustment for bidirectional QTL extension. Finally, a practical weakness of the application of considered RWR is that this procedure requires a great deal of thought, effort, and expertise, and there are many points in the procedure at which simple errors can produce dramatic changes in the confidence limits that are ultimately produced.

In light of the flaws of naive RWR, and the complexity of making the suggested adjustments, we recommend using the SPQV to assess the quality of QTL mapping experiments. The function provided, *SPQValidate*, requires only a few lists of data; the function itself accomplishes the analytic work that might be a stumbling block in RWR. Many of the other problems inherent to RWR are overcome by the SPQV's probabilistic nature. This tool is potentially overly conservative, however, in the case of short QTL. It is extremely unlikely that the SPQV will produce a value of 0 for the confidence limit, as any locus is likely to be within range of at least one marker for even the shortest identified QTL. Because of this, the minimum confidence limit is, in practice, 1, which might be misleading for small QTL. Additionally, the total number of genes in a QTL is not necessarily an authoritative measure of a QTL's validity; one can imagine that a QTL located on a single gene of high impact might be considered nonsignificant if the SPQV is the only method of validation used.

To use the SPQV, all QTL with a sufficiently high logarithm of the odds (LOD) score should be assessed. The function provides the higher and lower confidence limits, as well as the combined CI for each mathematically related set of QTL. If any QTL within a set exceeds the combined CI for that set, then the whole set can be deemed successful.

QTL with observed gene counts that exceed the upper confidence limits should of course be considered to have attained significance. If at least one QTL in an experiment is determined to be significant by the SPQV, the mapping experiment on a whole can be deemed a success. The QTL that contain lower numbers of genes might be identifying previously unknown genes, and therefore their non-significance does not detract from the significance of other QTL. Because of the potential for the identification of new phenotype-associated genes in QTL mapping, QTL containing significantly *low* numbers of identified genes might also be of interest, as they may have been placed on 'empty' regions due to previously unidentified genes that have a large impact on the phenotype. The QTL containing significantly low numbers of genes should only be considered interesting when at least one other QTL in the same mapping experiment has been proven to be significant so as to avoid the suggestion that all QTL in a faulty mapping study were significant.

CONCLUSION

This manifestation of the SPQV provides a tool to assess QTL one at a time, but it is often the case that more than one QTL is identified at a time. The current rendition of the SPQV acknowledges the multiplicity of QTL by combining the individual QTL confidence intervals. However, the order of QTL placement in a mapping experiment is important, as the first QTL placed will tend to occupy regions of the genome that are highly impactful in terms of the phenotype of interest, and are therefore more likely to identify known genes. Because QTL are restricted from overlapping, the first QTL restricts the genomic space onto which subsequent QTL can be placed, as well as the number of genes that it would be possible for the subsequent QTL to identify. The current version of the SPQV does not take this into account, and is therefore more stringent than necessary.

In all, the Scanning Probabilistic QTL Validator provides a strategy for QTL validation that takes into account the current understanding of the genes that contribute to the trait of interest. When used in conjunction with other methods like cross validation, the SPQV will allow researchers to identify QTL with potential for future applications with much greater certainty, and much less work, than has been previously possible.

Figure 2-1

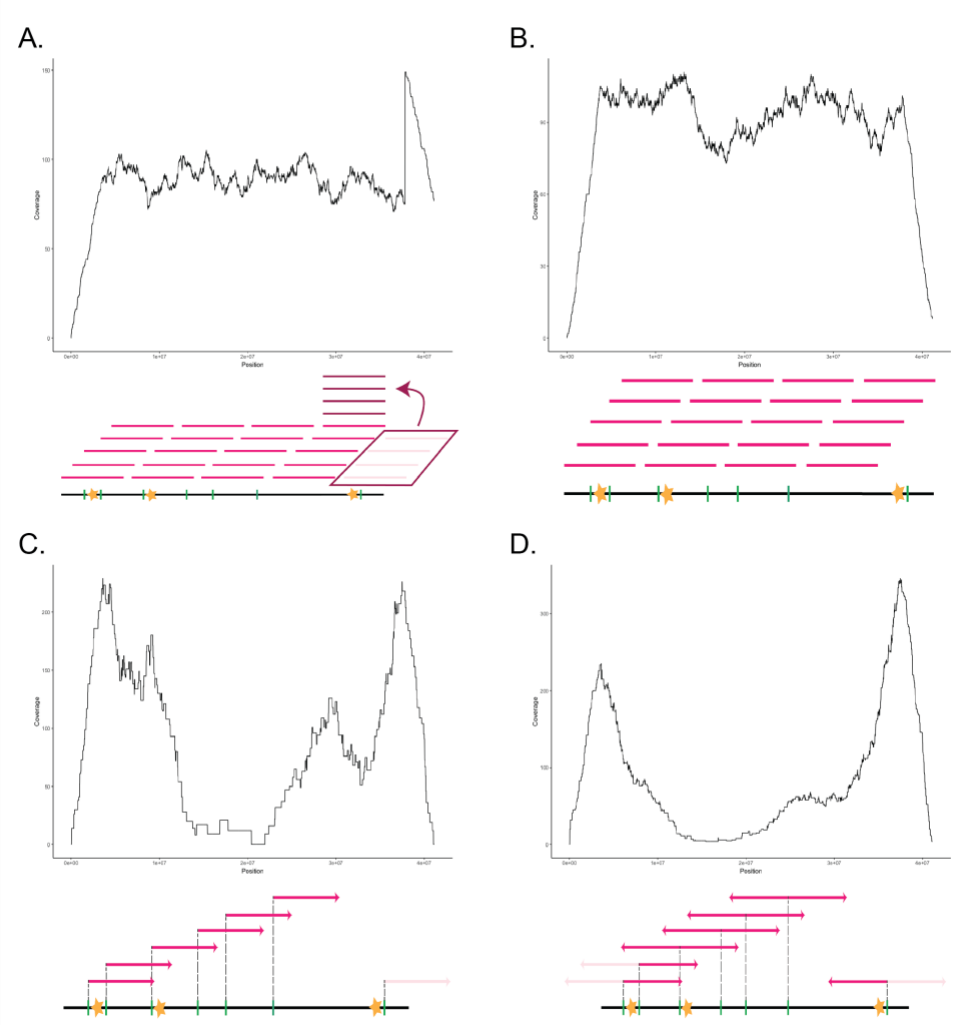


Figure 2-1. Effect on genome coverage of adjustments to RWR analysis of QTL mapping

The edges of chromosomes prove problematic for RWR analysis. The "bounceback" method (A) and the exclusion method (B) are two ways of handling problems associated with overhang. Both result in overrepresentation of regions that are unlikely to be mapped in practice due to the marker distribution on this chromosome. If RWR samples are constrained to begin on markers (C and D), the resulting coverage distribution is more reflective of the mapping process. Unidirectional QTL extension from a marker (C), though an improvement on the random placement of the origin shown in A and B, remains coarser than bidirectional QTL extension from a marker (D).

Figure 2-2

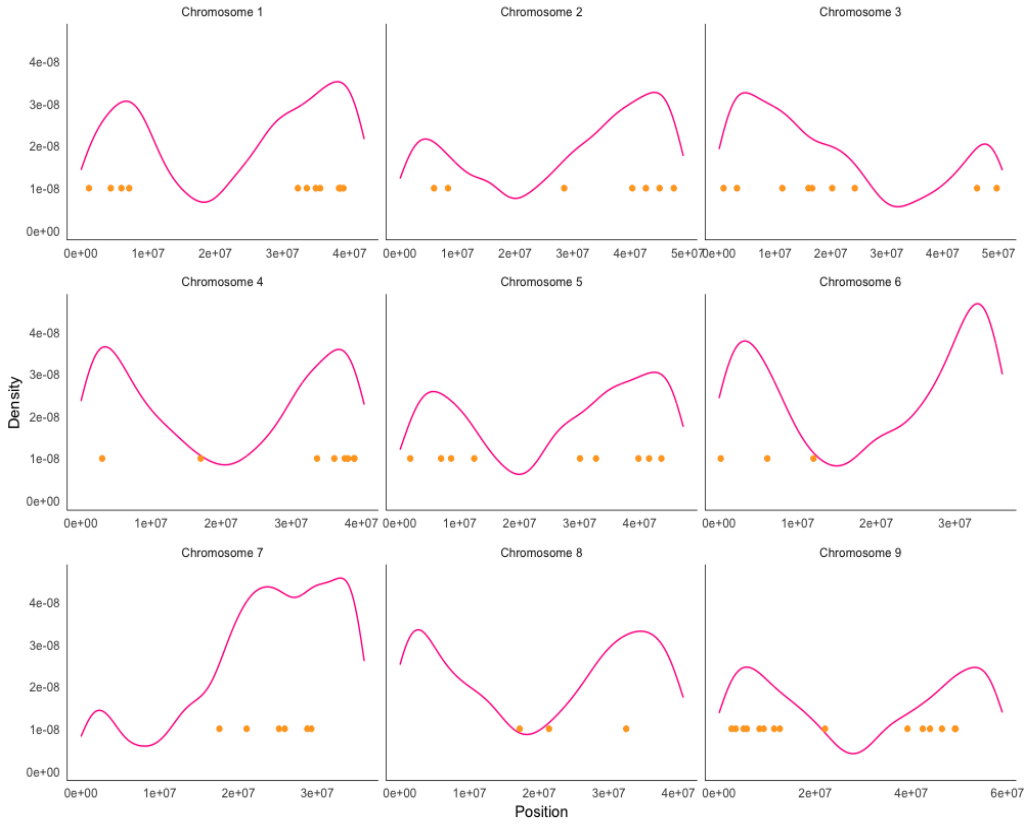


Figure 2-2. Whole genome gene density and single-trait gene density

The distribution of genes in *Setaria italica* is pictured; orange dots represent ion-associated genes. Chromosomal features such as telomeres and centromeres dictate much of the variation in gene density. While the distribution of known genes does not perfectly match the whole genome distribution, their similarities render the whole genome gene distribution a better choice for the production of simulated distributions than a random selection.

Figure 2-3

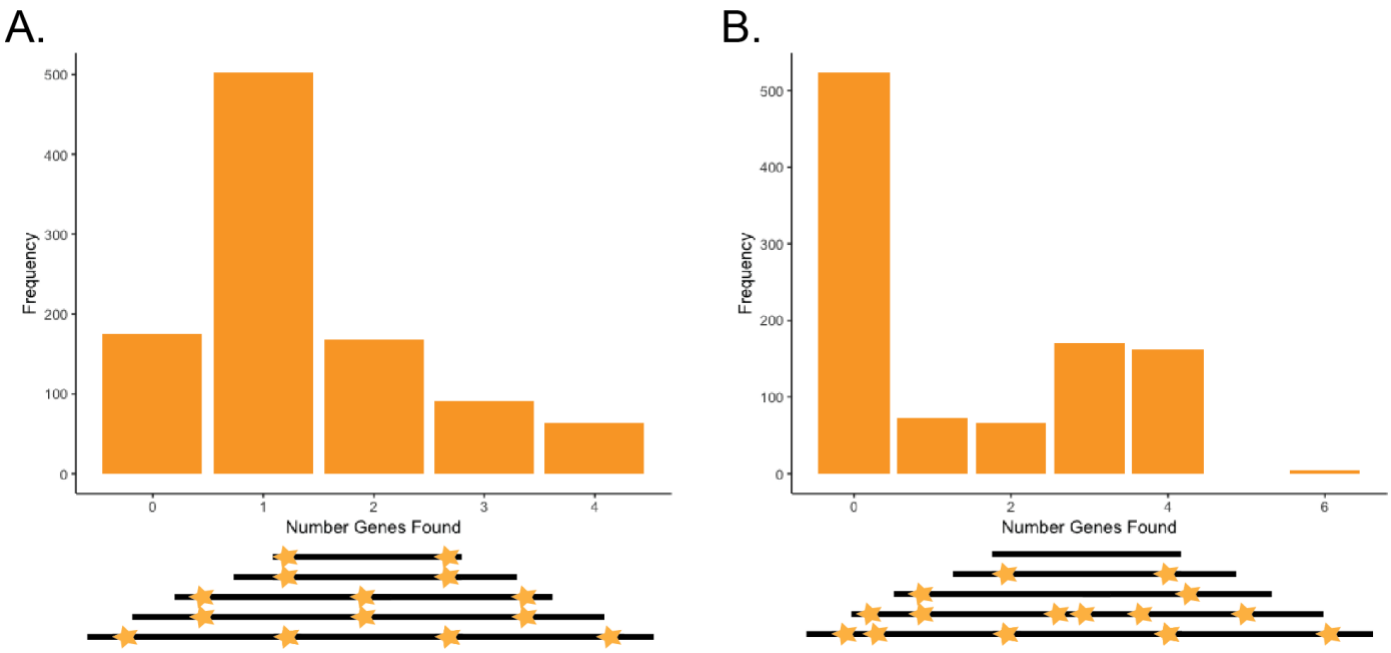


Figure 2-3. The effect of known gene density on RWR identified genes

If the density of known genes is consistent across chromosomes (A), the distribution of RWR identified genes approaches normality. However, genes associated with a phenotype typically have an uneven density distribution (B). In this case, the distribution of RWR identified genes tends to be unpredictably multimodal.

Figure 2-4

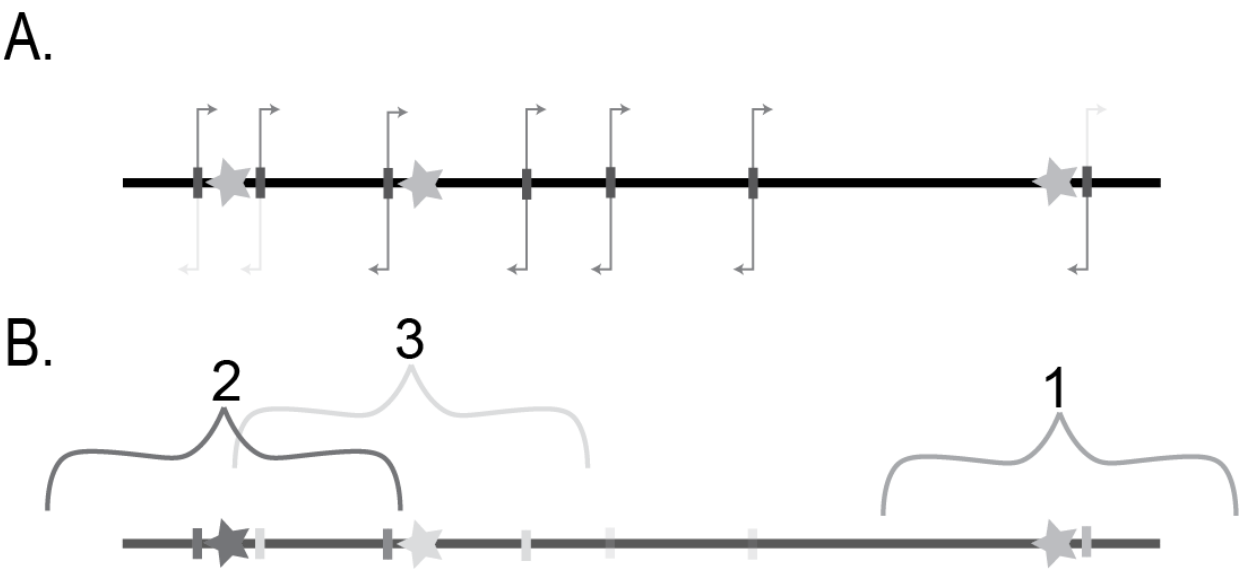


Figure 2-4. Calculating Expected Number of identified Genes (EGN)

The probability of selecting any marker in a given direction ($P(m_j)$) as the origin *O* for one RWR run is equal under the null hypothesis. Because QTL mapping is direction-independent, every marker is counted towards the sum of all possible *O* two times, with the exception of markers where the QTL would overhang the end of the chromosome in one or both directions (A). Light grey arrows indicate positions that are unavailable for a hypothetical QTL; darker grey arrows indicate possible positions for said QTL. The probability of identifying any locus ℓ_i with a QTL of length *L* is the sum of $P(m_j)$ for the markers within the range $[\ell_i-L, \ell_i+L]$. Markers can be counted twice if they can be used to map more than one locus. The sum of the probabilities for all loci produces the EGN. In this case, $P(m_j) = 1/11$, and the EGN is 6/11 (B).

Figure 2-5

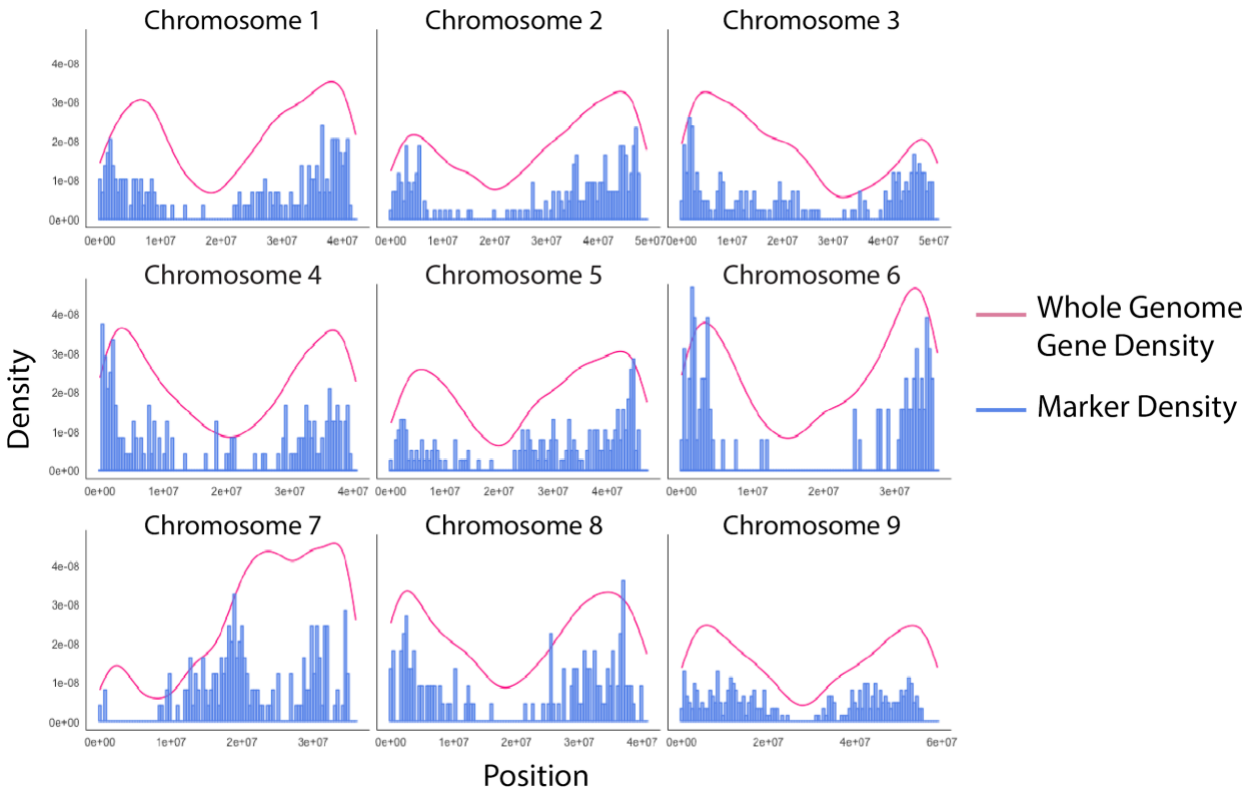


Figure 2-5. Marker placement corresponds to gene density in *S. italica*

The distribution of markers on all chromosomes reflects whole genome gene density in *Setaria italica* more closely than would a uniform distribution.

Figure 2-6

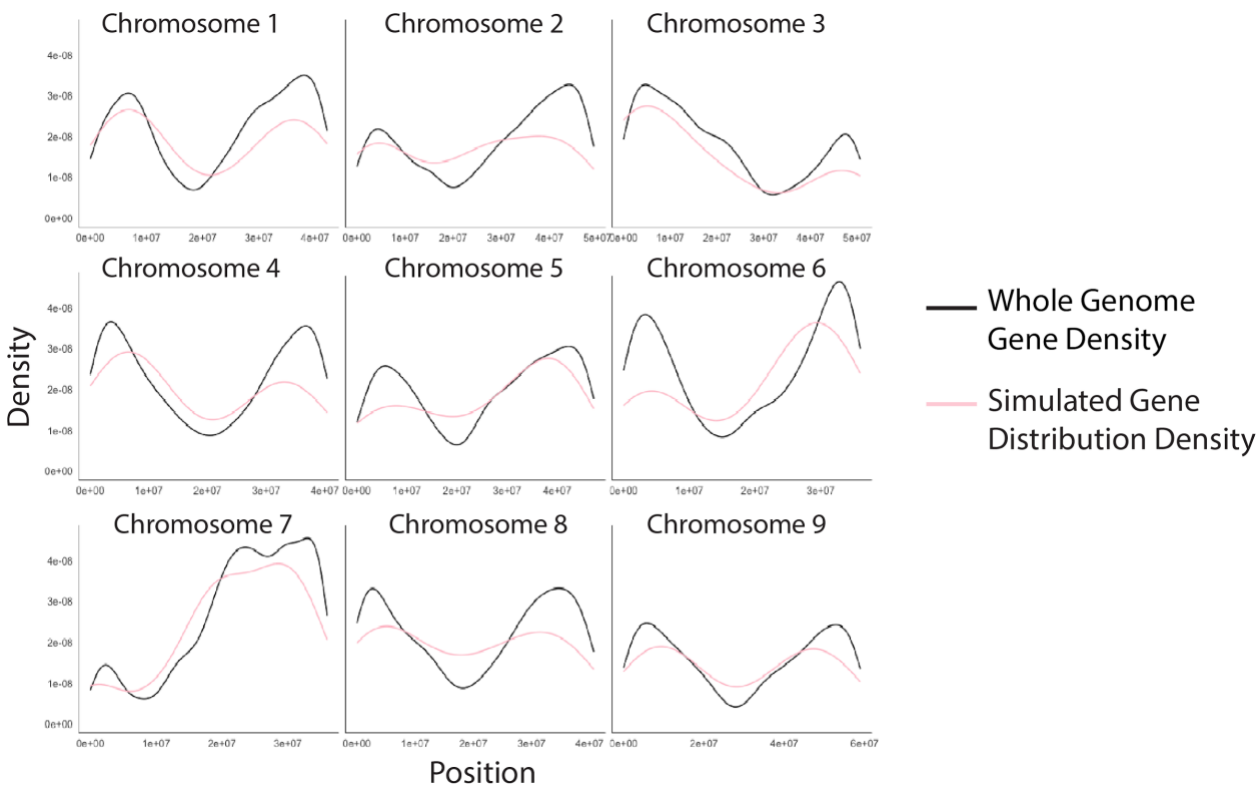


Figure 2-6. Randomly selected gene distributions reflect the genomic gene density

The simulated distribution of genes used to assay the RWR experiments in Figure 2-7 reflects the whole genome gene distribution.

Equation 2-1

$$\mathrm{E[Gene\ Number]} = \sum_{i=1}^{G} \sum_{j\ in\ range\ of\ \ell_i}^{M_i} P(m_j)$$

Equation 2-1. Calculating expected number of identified genes

G is defined as the number of known genes in the genome, M_i is defined as the number of markers in range of a locus ℓ_i, and $P(m_j)$ is defined as the probability of selecting any individual marker for mapping purposes in the mapping direction that would cause a QTL of given length L to overlap the locus at ℓ_i.

Figure 2-7

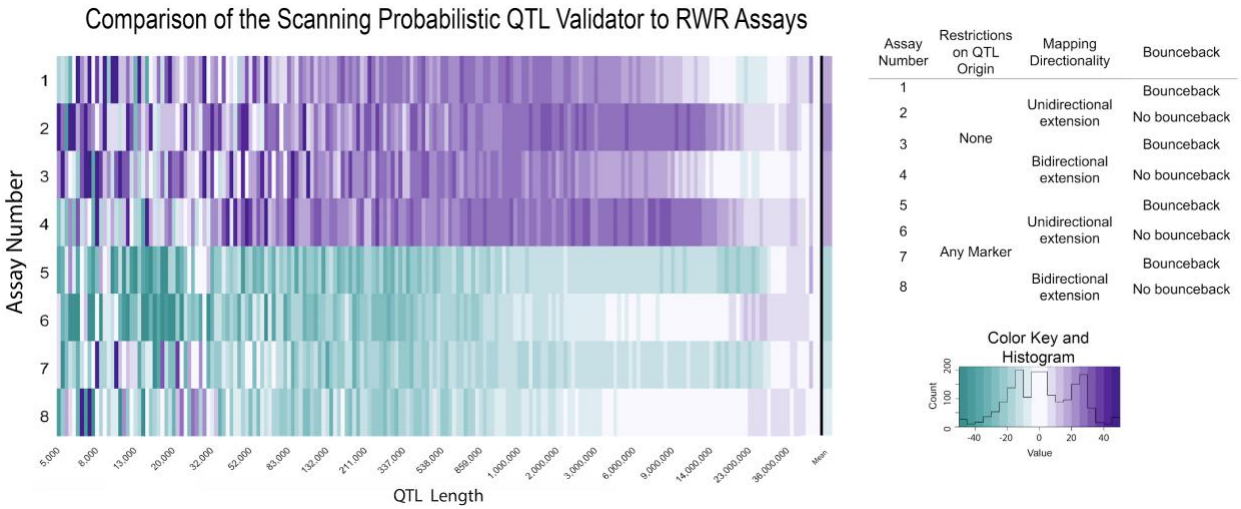


Figure 2-7. Comparing the Scanning Probabilistic QTL Validator with RWR Assays

The SPQV was used to assess 199 hypothetical QTL on a simulated trait dataset of 333 genes that were randomly selected from the genome, all of which are shown here in order of increasing QTL size. The final column on the right reflects the mean difference between the SPQV and the RWR assay for all QTL sizes. The color indicates the percent difference ((x - y) / mean(x, y)) between the upper confidence limits produced by various RWR assays and those resulting from the SPQV. Values above 0 indicate that the SPQV value is greater than the RWR estimated values, and values below 0 indicate the reverse. The SPQV consistently identifies confidence limits that are approximately in line with the RWR assay that takes the most of the various biologically relevant factors into account (row 8).

Figure 2-8

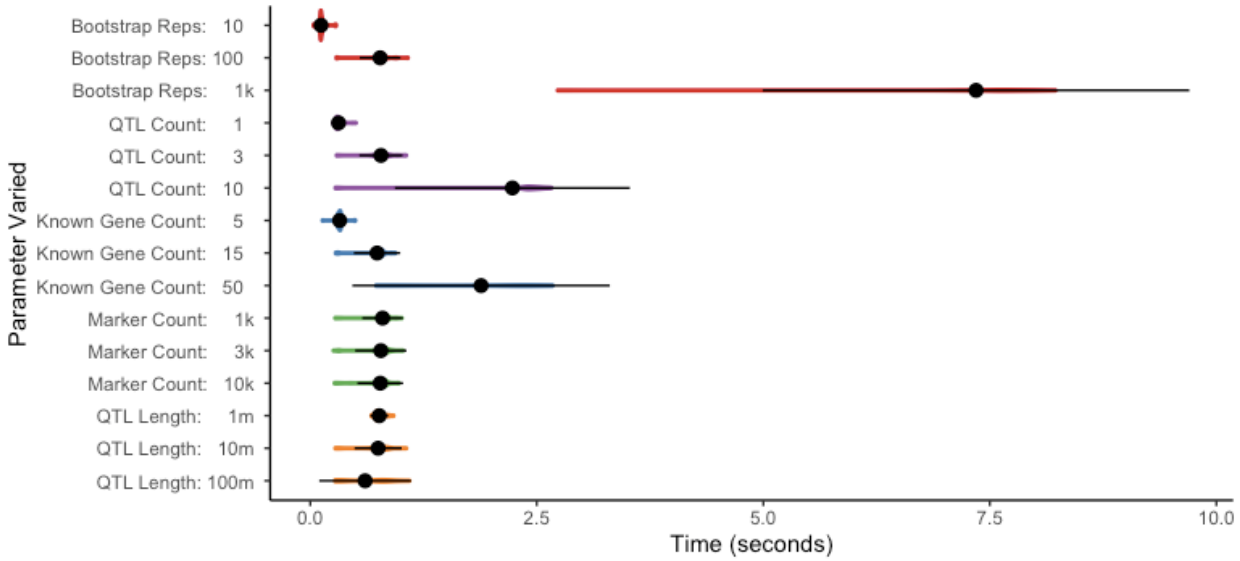


Figure 2-8. Runtime analysis of provided SPQValidate() function

The function was run using varying data sizes and RWR repetitions. Each experiment was run 100 times, and the distribution of time taken per run is shown. Only one parameter was varied at a time; for the remaining parameters, the median of the three values shown was used. The script used can be found in the "paper_figures" section of our GitHub repository.

Table 2-1

GeneID	Chromosome	Locus midpoint
TB1	1	272844092
GT1	1	24332633.5
TRU1	3	154277952
TGA1	4	45787466.5
TSH4	7	136485942
CG1	3	7747826
HSF1	5	209689354

Table 2-1. Genes associated with tillering in maize

Genes known to be associated with tillering in maize were identified in the Zm-W22 NRGene 2.0 assembly (Springer *et al.* 2018). Each gene was selected based on the strength of the association with tillering; more weakly associated genes could lead to an overly stringent implementation of the SPQV.

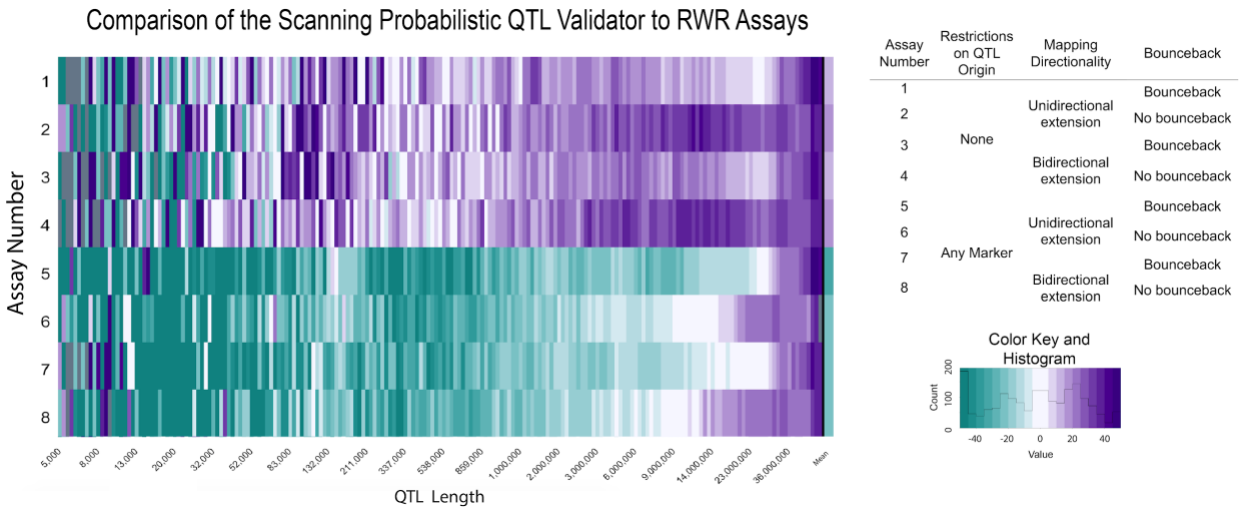
Table 2-2

Trait	QTL Length	Observed Value	Mean EGN	SEM	Lower 95% CI	Upper 95% CI	Additive Lower 95% CI	Additive Upper 95% CI
KRN	161,766,731	**1**	0.427	0.145	0.142	**0.712**	0.170	**0.741**
	3,783,574	0	0.017	0.006	0.005	0.029		
	1,904,718	0	0.009	0.003	0.002	0.015		
	550,190	0	0.003	0.001	0.000	0.005		
TILN	83,615,820	0	0.239	0.053	0.136	0.342	0.256	**0.485**
	20,375,560	0	0.079	0.021	0.038	0.119		
	13,021,112	**1**	0.053	0.015	0.023	**0.083**		
EB	84,725,594	**1**	0.242	0.054	0.137	**0.347**	0.137	**0.347**
STAM	945,639	**1**	0.004	0.002	0.001	**0.008**	0.001	**0.008**
BARE	6,392,742	0	0.028	0.009	0.009	0.046	0.009	0.046
GLUM	1,694,383	0	0.008	0.003	0.002	0.014	0.002	0.014

Table 2-2. Results of SPQV assay of branching QTL identified in the TeoNAM population

Sums for multi-QTL trait groups represent additive confidence intervals constructed by summing the mean and variance of the EGN for each QTL. Values that achieved significance are bolded.

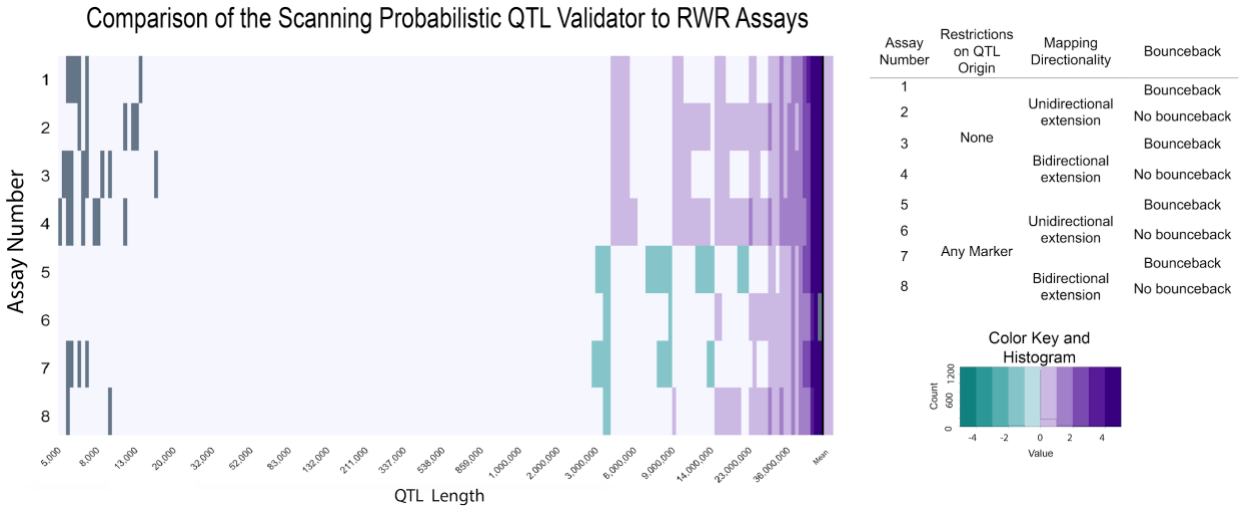
Supplemental Figure 2-1



Supplemental Figure 2-1. Comparing the Scanning Probabilistic QTL Validator with RWR Assays using a trait associated gene list

71 genes known to be associated with the ionome in *Setaria italica* were used as the test distribution. The same hypothetical QTL were used as in Figure 2-7. In this case, RWR of several of the smallest QTL produced distributions that were not able to produce BCa confidence limits. These instances are indicated in grey. The reduction in size of the list of known genes produced confidence intervals that were more in line with RWR in all assays, but similar patterns to those in Figure 2-7 are observable.

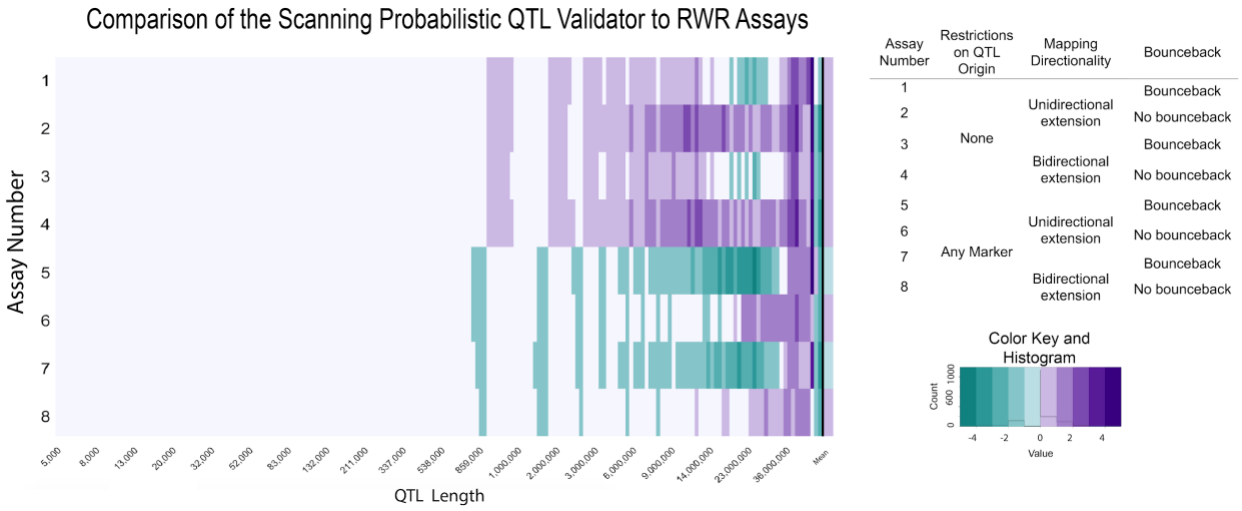
Supplemental Figure 2-2



Supplemental Figure 2-2. A ceiling comparison of the Scanning Probabilistic QTL Validator with RWR assays using a trait associated gene list

The data used to produce Supplemental Figure 2-1 were rounded up to the nearest whole number to illustrate the differences in working confidence limits between the SPQV and RWR. The grey bars that occur in the columns of the smallest QTL represent NA values.

Supplemental Figure 2-3



Supplemental Figure 2-3. A ceiling comparison of the Scanning Probabilistic QTL Validator with RWR Assays

The data used to produce Figure 2-7 were rounded up to the nearest whole number to illustrate the differences in working confidence limits between the SPQV and RWR. The smaller QTL have essentially the same values regardless of method.

Chapter 3. Interactions between genotype and water status drive variation in the *S. italica* ionome

INTRODUCTION

The intake, transportation, and storage of elements within a plant is vital to its appropriate growth and development. Plants must maintain balance between excessive uptake, which can cause toxicity and necrosis through the production of free radicals or the exclusion of other nutrients, and sufficient uptake of essential inorganic nutrients (Bollard and Butler 1966; Brown and Jones 1975; Rout and Das 2009). Tight control of ion homeostasis allows plants to respond appropriately to environmental conditions such as temperature, soil pH, and water availability (Wang *et al.* 2003). This regulatory control must be responsive to both the concentration of ions within the soil and to those concentrations within the plant. Understanding the genetic control of elemental uptake and transportation will contribute to efforts to improve crops, and to the ultimate sustenance of the growing global population.

Examining the concentration of elements within a plant sample allows for exploration of the genetic and physiological processes involved in adaptation to a particular environment. To this end, we and others have developed a pipeline that is designed to cheaply and efficiently measure the concentrations of 20 different elements via inductively coupled mass spectrometry (ICP-MS) (Salt *et al.* 2008; Feldman *et al.* 2017). This process is known as ionomics, which is defined as the quantitative study of the mineral nutrient and trace elemental content of an organism; that is to say, its ionome (Salt *et al.* 2008). In cereal crops such as *Setaria italica*, the flag leaf emerges just before the panicle, and therefore marks the specific developmental time point at which the plant has taken up the majority of its total mass. This tissue is therefore ideal for ionomic inquiry. Additionally, nutrient loading of the grain is accomplished through the remobilization of organic and inorganic materials from the leaves (Kichey *et al.* 2007; Gregersen *et al.* 2008).

The flag leaf specifically is instrumental in the loading of photoassimilates and other micronutrients (Gregerson *et al.* 2008; Jeong *et al.* 2017; Zhang *et al.* 2006), and is therefore commonly used as representative tissue to assess the composition of a grass (Evans 1983; Cuin *et al.* 2008; Carey *et al.* 2011).

The species *Setaria viridis*, or green foxtail millet, is a member of the Panicoideae that utilizes C4 photosynthesis (Brutnell *et al.* 2010). It is therefore a good model system for several related, economically important crops, including sorghum and maize. The compact stature, short life span, and sequenced genome of *S. viridis* have also contributed to its status as an emerging model organism (Brutnell *et al.* 2010; Bennetzen *et al.* 2012). In addition, *S. viridis* is the wild ancestor of the crop species *Setaria italica* (foxtail millet); these two species have a semi-permeable boundary between them, as their primary difference is phenotypic and they are still readily crossed (Dekker 2004).

Foxtail millet is a member of the small millet species, a group of ancient grains that are relatively nutritionally dense when compared to rice and wheat and which are often cultivated as subsistence crops (Ravindran 1991; Goron and Raizada 2015). Because of the combination of the high nutritional value of the *S. italica* grain and its resistance to abiotic and biotic stressors, breeding elite cultivars of *S. italica* for increased nutritional content is an attractive prospect. Understanding the contribution that different regions of the *S. italica* genome make to total nutrient content is an important first step for breeding purposes.

Structured populations are useful tools for dissecting the relationship between elemental accumulation and the genetic content of a species. Recombinant inbred lines (RILs) have various advantages when it comes to quantitative trait loci (QTL) analyses. Repetitive selfing allows for the break up of large linkage blocks, which in turn allows for finer mapping (Takuno *et al.* 2012). Additionally, once established, RILs may be continuously maintained in a fixed homozygous state. This makes it possible to assay the same combinations of alleles in multiple different environments (Takuno *et al.* 2012). The resulting phenotypic and genotypic data can then be compared through various statistical means in order to identify QTL.

Here, we use elemental profiling of a RIL population resulting from a wide cross between *S. italica* and *S. viridis* grown in multiple environmental conditions to identify QTL associated with the ionic content of leaf material. Overall, we identified 251 QTL, 171 of which were associated with a single element and 80 of which were associated with a principal components analysis of the ionome. The use of traits defined by the differences between treatments in an experiment allowed for the quantification of the influence of the environment on *Setaria's* ionome.

MATERIALS AND METHODS

Population, Field Growth, and Data Collection

Experiments were conducted in the summers of 2013 and 2014. Experiments assaying the effect of density of planting on ionic content were conducted in 2013 and 2014. A single drought experiment was conducted in 2014. A total of 189 F7 RILs resulting from a wide cross between the B100 cultivar of *S. italica* and the A10 line of *S. viridis*, together with their parent lines, were used as the study material (Bennetzen 2012). In every experiment, lines were planted in triplicate in a block design in the field in Creve Coeur, MO. Treatments in the density experiment consisted of either five centimeter spacing between neighboring plants, or twenty centimeter spacing between neighbors. Plants in the drought experiment were either well watered until the time of sample collection, or were subjected to drought stress from eight weeks post planting. Flag leaves were harvested from plants four days after panicle emergence.

Elemental Profile Analysis by ICP-MS

The Baxter laboratory has a documented, standardized pipeline for elemental analysis (Ziegler *et al.* 2013), which is described in detail in Chapter One of this thesis. Notably, for this chapter the measurements for the elements B, Mn, S, and Se were not included in these analyses, as the concentration of these elements did not depend appreciably on genotype.

Computational Analysis

Outlier removal and variance partitioning: Elements were treated individually within each dataset. For each element, a first pass of analytical outliers were removed as in Davies and Gather 1993. Briefly, data points were considered global outliers if their median absolute deviation (MAD) was greater than 6.2 as calculated within line and experiment.

In order to determine if genotype contributed to the elemental phenotypes measured to a degree sufficient for mapping, the variance in the data was partitioned using the lmer function from the lme4 package in R (Bates *et al.* 2015). Variables including genotype, plot, treatment, and the interaction between treatment and genotype were assessed for their contribution to the total variance in the datasets. To do this, data lacking the global outliers mentioned above was modeled using two mixed linear models. The first was as follows:

$$Y_i = g_i + t + p_i + g_it$$

in which Y_i indicates the phenotype of the individual i, g_i indicates that individual's genotype, t indicates the treatment applied, and p_i represents the plot in which the plant was grown. g_i , p_i , and g_it were treated as random effect variables, while t was treated as a fixed effect variable. The second model lacked the term t, but was otherwise identical to the first. The variance was extracted from each term in each equation using the function *VarCorr* from the package nlme (Pinheiro *et al.* 2019). Because this program cannot attribute variance to fixed effect variables, the variance for t was calculated by subtracting the total variance that was attributed to any variable in the first equation from the total variance attributed in the second equation.

The data were then transformed to normality using the Box-Cox family of transformations. The specific transformation for each element in each experiment was determined by the default R function boxcox() and are listed in Supp. Table 1. To identify and remove any remaining significant outliers in the transformed data, a simple mixed linear model was fitted for each element using ASReml-R version 3.0 (Gilmour et al., 2010). The model fitted to the data was:

$$Y_i = g_i + t + p_i$$

This model was selected due to the combination of its simplicity and its inclusion of the variables that had large contributions to the variation present in the data. Genotype and plot were treated as random effects variables, while treatment was treated as a fixed effect variable. Studentized deleted residuals (Neter et al. 1996) were calculated based on the above model and subsequently used to detect any remaining outliers.

Heritability Calculations: Broad-sense heritability was calculated for each measured element within each of the three experiments. To do this, the total phenotypic variation was partitioned into genetic and environmental variance, as well as variance attributable to the interaction of genotype and treatment. The broad sense heritability corresponded to the genetic portion. A model was fit to the data using the *lmer* function from the lme4 package in R (Bates *et al.* 2015). This model was as follows:

$$Y_i = g_i + t + p_i + g_i t$$

Each of these terms was treated as a random effects variable, with the exception of *t*. The variances for genetics, environment, and GxT were extracted using the function *VarCorr*. The broad sense heritability was subsequently calculated using the formula

$$H = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_{GxT}}{hm_T} + \frac{\sigma^2_\varepsilon}{hm_{tot}}}$$

The variable hm_T was calculated as the harmonic mean of replication within treatment blocks and hm_{tot} was calculated as the harmonic mean of experiment wide replication. Only RILs with greater than 2 replicates within each experiment were used for the calculation of heritability. Global outliers, as identified through the MAD method described above, were designated as "NA" for these calculations.

Finally, repeatability was calculated for each element across the three experiments. A simple model, $Y_i = g_i$, was fit to the data as above, and the variance for genotype was similarly extracted. The repeatability was calculated using the equation

$$R = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_\varepsilon}{n}}$$

where *n* is defined as the number of observations, in this case experiments. All calculations were also performed on ranked data.

BLUP development and QTL mapping: A mixed model for each element in each location was selected through an iterative process using ASReml-R version 3.0 (Gilmour et al., 2010). The models' goodness of fit were assessed using the Akaike Information Criteria (AIC). The equation

$$Y_i = g_i + t + p_i$$

in which genotype and plot were treated as random effects variables and treatment was used as a fixed effect variable, was invariably identified as the best fit model. The final model was used to generate Best Linear Unbiased Predictions (BLUPs) for each element within each experiment.

For each element within an environment, these values were formatted into an R/qtl cross object along with the line average phenotypes, genotypes for each line, and 1595 biallelic single nucleotide polymorphisms (SNPs) that resulted from a GBS derived genomic map (Feldman *et al.* 2017). QTL mapping was performed as in Feldman *et al.* 2017.

In sum, five phenotypic metrics for each ion were used for mapping. Mapping was performed both within treatment blocks as well as on the relative difference, numerical difference, and the trait ratio (defined as the ratio of values from different treatment blocks). The functions used were taken from the R/qtl and R/funqtl packages (Broman et al., 2003; Kwak et al., 2016), and called using the custom R and python scripts in the foxyqtlpipeline (Feldman *et al.* 2017). Additional mapping was performed on the rotated loadings resulting from principal components analyses of the BLUP modeling results for each individual experiment using the native R function prcomp. Only principal components explaining greater than two percent of the variation present in the data were retained. Elements that were poorly measured were removed prior to PCA in order to avoid skewing the calculations.

Two separate but complementary strategies were used to map the data. Firstly, QTL with logarithm of odds (LOD) score peaks exceeding a significance threshold based upon 1000 permutations (a= 0.05) were identified using a genome scan under a single QTL model as calculated by a Haley-Knott regression. Secondly, a multiple QTL model was produced using a stepwise forward/backwards selection process. This latter model used a penalized LOD score criterion. QTL associated with the differential metrics described above may identify loci associated with genotype by environment interactions (Des Marais et al., 2013).

QTL Validation: The positions of the QTL were validated using the Scanning Probabilistic QTL Validator (SPQV) (Haining and Blumer, Chapter Two).

This tool leverages the current knowledge about the genes that contribute to a trait ('known genes') in order to determine if a QTL has identified a statistically significant number of known genes. The metadata associated with all of the identified genes in the whole genome of *S. italica* was extracted from the Ensembl Plants website (Kersey *et al.* 2018) and used to produce the whole genome gene distribution. This distribution, the markers used for mapping, and a list of 147 ionomic genes inferred from other species were input into the R function *SPQV*. For assaying ion-specific QTL, only the genes associated with that ion were used for validation; for QTL resulting from principal components analysis, all 147 genes were used. If both confidence intervals of an identified QTL had LOD scores greater than 1, that QTL was analysed for significance.

RESULTS

The data used for this work included measurements for 20 different elements in flag leaf tissue collected from a recombinant inbred line (RIL) population resulting from the cross of the B100 cultivar of *Setaria italica* and the A10 line of *Setaria viridis*. A drought experiment was conducted in Creve Coeur, Missouri in 2013 (DN13); both a drought and a density experiment were conducted in the same location 2014 (DN14, DR14). 179 of the RILs were planted in at least two of the three experiments, while 116 were grown in all three. The leaf samples from all experiments were treated in an identical fashion; samples were dried and stored in temperature and humidity controlled rooms before ionomic analysis.

Each sample was profiled for the quantity of 20 elements using ICP–MS (Figure 3-1). The resultant measurements were normalized to the sample weight and technical sources of variation using a linear model. Experiment level analytical outliers were removed as in Davies and Gather 1993. Pursuant to this, the measured values for each element were transformed to normality using the Box Cox family of transformations, and Studentized deleted residuals (SDR) were used to identify and eliminate further outliers within the measurements for each element. After outlier removal, phenotypes were derived by averaging the values for each line within an experiment and treatment.

Both environment and genotype impacted the variation present in these data. Repeatability was generally lower than within experiment heritability (Table 3-1), indicating that there was less variation in genotypic replicates within an individual experiment than across experiments. The broad sense heritability (H^2) of 9 elements in the DN13 experiment, 14 elements in the DN14 experiment and 16 elements in the DR14 experiment exceeded 0.4. Certain elements including selenium, sulfur, and boron showed low repeatability; this is likely due to the fact that these elements tend towards analytical artifacts, as they accumulate to levels that are near the limits of detection of the methods described in this paper. The heritability of individual elements varied by up to 0.533 between different experiments.

The function *stepwiseqtl* from the R package was used in order to identify a multiple QTL model for each of the elemental phenotypes. This function moves iteratively through the genome to test for significant allelic effects of each marker on the phenotype in question. When a significant locus has been identified, this is added to the model. A combination of forward and backward regression ultimately produces a genome wide QTL model for each trait. Each element was considered individually, as well as in combination with the others as a contributor to a principal components analysis that was run for each experiment (DN13, DN14, and DR14). Five different metrics were used as the phenotype for each element in each experiment. These phenotypes include the 'raw' values for each treatment (e.g. well watered, droughted) and the differential values for that trait (relative difference between treatments, difference between treatments, and ratio of the treatments). The significance of a QTL was computed using the 95th percentile threshold resulting from 1000 iterations of the scanone function as a penalty for adding the QTL to the model.

When all experiments are considered, a total of 251 QTL were identified (Table 3-2). As expected from the heritability measurements, the majority of these were identified in the 2014 drought experiment (105 QTL). The 2013 and 2014 density experiments allowed for the identification of 75 and 71 QTL, respectively. Approximately a third of the QTL (81 of 251; Figure 3-2 A) were identified within treatments; the remainder were identified using either the difference, relative difference, or ratio of the phenotypic values measured within the different treatments in a single experiment (Figure 3-2 B). Of the 251 QTL, 80 were identified for the mapping based on the principal components analysis; 39 of these resulted from the drought experiment, 21 from the 2013 density experiment, and the remaining 20 from the 2014 density experiment. Of the 251 QTL, 55 were located on chromosome 2 (Figure 3-3; Figure 3-4).

The locations of the QTL were assessed for overlap with the locations of known ionic genes (Supp. Table 3-2). Forty five of the QTL contained at least one gene within their 95% confidence intervals. Of the QTL that coincided with genes, 35 were identified using PC as the mapped trait. The QTL were assessed for significance using the Scanning Probabilistic QTL Validator. The QTL were divided into mathematically related sets based on experiment and the phenotypic metric that was used when they were mapped. The results of this assessment are reported in Table 3-3; each phenotypic metric that was used for mapping was associated with at least one set of QTL that identified a significant number of genes, indicating that the data curation was done effectively.

There were several regions in which QTL were remarkably concentrated. Fifteen QTL were identified on chromosome 2 between 89.4 and 95.9 cM (Figure 3-4); these QTL were discovered in both the 2013 density experiment and the 2014 drought experiment. The traits associated with these QTL included As, Al, Co, Cu, Mo, P, Rb, and Sr, as well as PC2 for the 2013 density experiment.

A similar stack of QTL occurred on chromosome 5 at 111.9 cM; associated traits included Ca, Cd, Sr and PC1 for the 2013 density experiment. In the 5.4 cm between position 109.7 and 115.1 cM on chromosome 5, 6 QTL were identified, with the additional 2 QTL being associated with Sr (Figure 3-5). Another group of QTL was found on chromosome 9 at 123.7 cM which contained QTL for Ca, Ni, Sr, and PC9, all from the drought experiment. Finally, 4 QTL were found on chromosome 7 at 99.9 cM associated with P, PC1 for DN14, and PC6 in DN13. When the range for this stack was extended to chromosome 7 at 95.3 cM, two more QTL were found, both of which were associated with As (Figure 3-5).

Of the QTL that were identified for specific ions, 29 were found using at least two phenotype metrics, 3 in three or more phenotype metrics, and one QTL for Zn concentration was identified using four phenotype metrics. Nineteen of the 80 PC associated QTL were found in at least two separate experiments, with eleven of these identified in both the drought experiment and at least one density experiment.

One gene, Sevir.5G106900, was identified by four separate ion specific QTL (Table 3-3) and five QTL based from data from the principal components analysis, including PC1 from DN13 and DR14 and PC2 from DN13. A neighboring gene, Sevir.5G144600, was found in the same QTL for PC1, but only in a single As specific QTL.

DISCUSSION

In the context of the ionome, principal components analysis allows for the identification of regions of the genome that would not otherwise be found. While some of the PC QTL identified regions that overlapped with those identified by ion specific QTL, the majority of them, including many QTL identified for PCs which explained a large amount of the variance present in the data, did not. It is possible that for a single ion, the signal associated with the PC QTL regions is not sufficient for their identification, while the additive signal that is inherent in a principal component suffices. Moreover, many of the first few principal components overlay regions associated with water use efficiency (WUE), with concentrated regions of QTL identified on chromosomes 2, 5, 7, and 9 at positions 94, 111.9, 99.9, and 123.7, respectively. A similar pattern was observed in Feldman *et al.* 2017, in which the authors identified QTL on chromosome 2 at 96cM, 5 at 109cM, 7 at 99cM, and 9 at 127 cMwhich were associated with water use efficiency (WUE) using the same A10 x B100 RIL population (Figure 3-6). Though the processes of water and ion uptake are independent, there is a strong relationship between the two; ion homeostasis depends upon transpiration rate, active transport, and membrane permeability, all of which are affected by the water status of the plant. As drought and density of planting are both variables that impact the available water supply, it appears that alteration in the water status of *Setaria* substantially perturbs the ionome.

This additionally suggests that, while the ionome can be interrogated with individual ions, a multi-elemental approach is more likely to identify regions of the genome with weak signal, or those that evince pleiotropy.

A large fraction of the QTL were identified for differential traits (Figure 3-4). Rb is a striking example of this, with 13 of its 15 QTL identified via differential analysis. Combined with this trait's high heritability within each experiment and low repeatability across experiments, these data suggest that there is a strong genotype by treatment component in Rb content in this RIL population. This genotype by treatment effect was apparent in many elements that were assayed, with an average of 74% of the ion specific QTL identified in the differential traits. The preponderance of differential QTL was not universal, as within treatment mapping allowed for the identification of 50% or more of the QTL found for Mo and Sr, suggesting that the homeostasis of these elements experiences a smaller degree of environmental perturbation.

Several QTL identified in this study overlie genes known to be associated with the control of elemental concentration. One example is Sevir.5G251200, the ortholog of MOT2 in *A. thaliana*. MOT1, a MOT2 paralog, is responsible for a large fraction of the variation in Mo content in *A. thaliana* (Baxter *et al.* 2008). Although MOT1 has an ortholog in *S. viridis* on chromosome 9, it was not identified by any of the QTL in this study. This finding suggests that either 1) the RIL population contains allelic diversity in MOT2 that is not present at the chromosome 9 locus or 2) the MOT2 locus in *Setaria* is responsible for more of the variation in Mo content in this species. Additionally Sevir.5G106900, imputed from the *A. thaliana* gene ESB1, underlay 10 QTL, the majority of which were identified in the PC QTL mapping. ESB1 is involved in the production of the casparian strip, with mutants in *A. thaliana* developing increased suberin levels and disordered casparian strips, as well as altered levels of many ions. The identification of QTL in this region in both the DN13 and DR14 experiments and for several different treatments, as well as in the first principal component for both DN13 and DR14 suggests that the *Setaria* ESB1 ortholog plays a role in a variety of conditions related to water status. The B100 haplotype for this region produces a decrease in Mo as compared to the A10 allele, which is consistent with the relationship between the WT and the *esb1* allele seen in *A. thaliana*. Given the central role played by the casparian strip in water homeostasis, this gene is a good candidate for explaining the coincidence of WUE and ionomic QTL.

CONCLUSION

In this study, we have shown that the ionome of the *Setaria* flag leaf depends upon environmental and genotypic factors, with a strong component attributable to the interaction between these influences. Assaying the differences between treatments within experiments allowed us to identify the regions of the genome associated with G x E.

The use of the SPQV as outlined in Chapter 2 of this thesis illustrated the validity of the mapping results. Additionally, the use of principal components analysis revealed that regions associated with WUE in this population also influence the ionome. For best effect, future work should entail focus on the analysis of *Setaria*'s ionome in the context of multiple treatments.

Figure 3-1

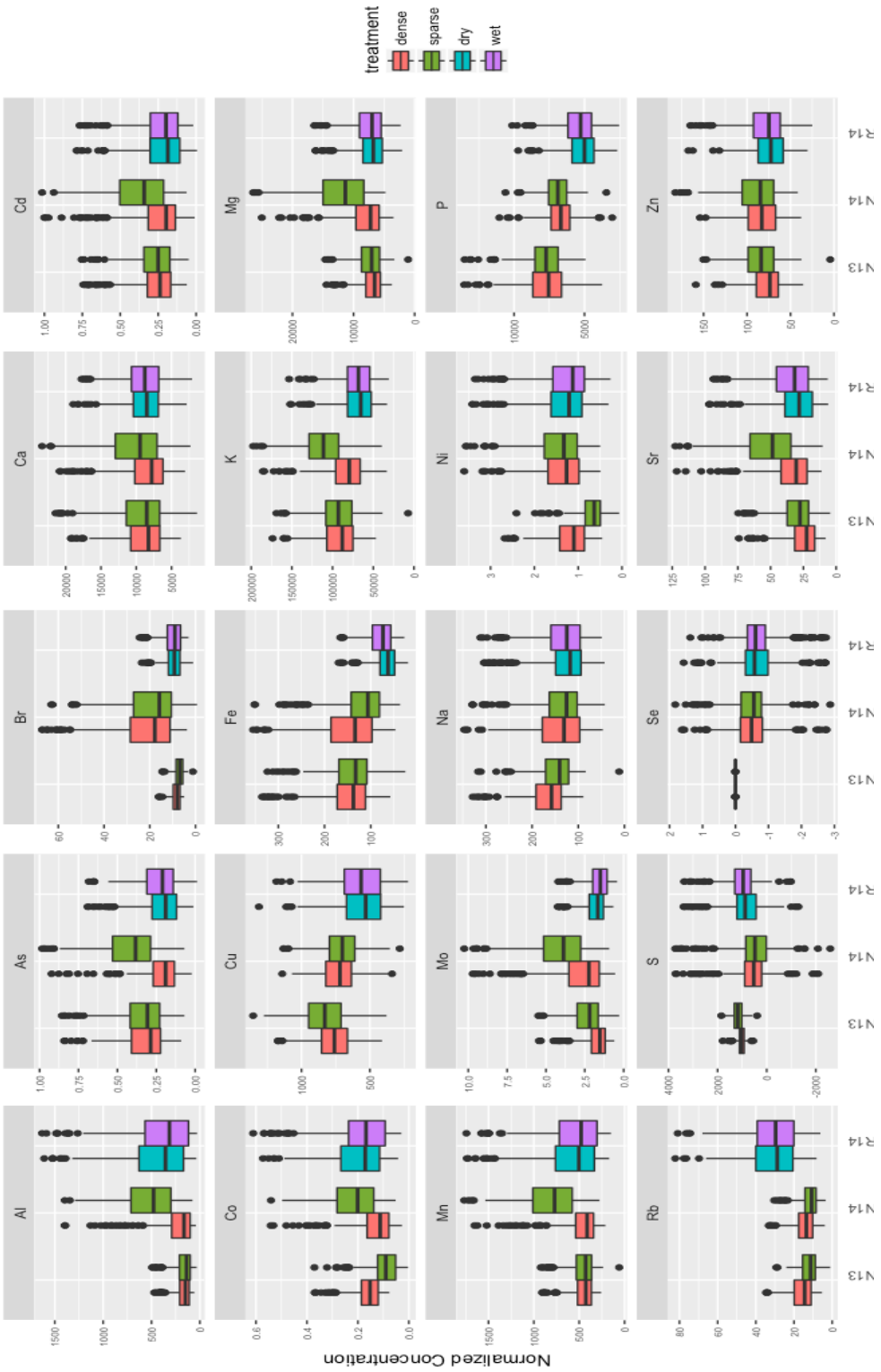


Figure 3-1. Normalized concentration of 20 elements in 4 treatments in the *Setaria* flag leaf

Data were plotted before outlier removal to illustrate the full range of values that were collected.

Figure 3-2

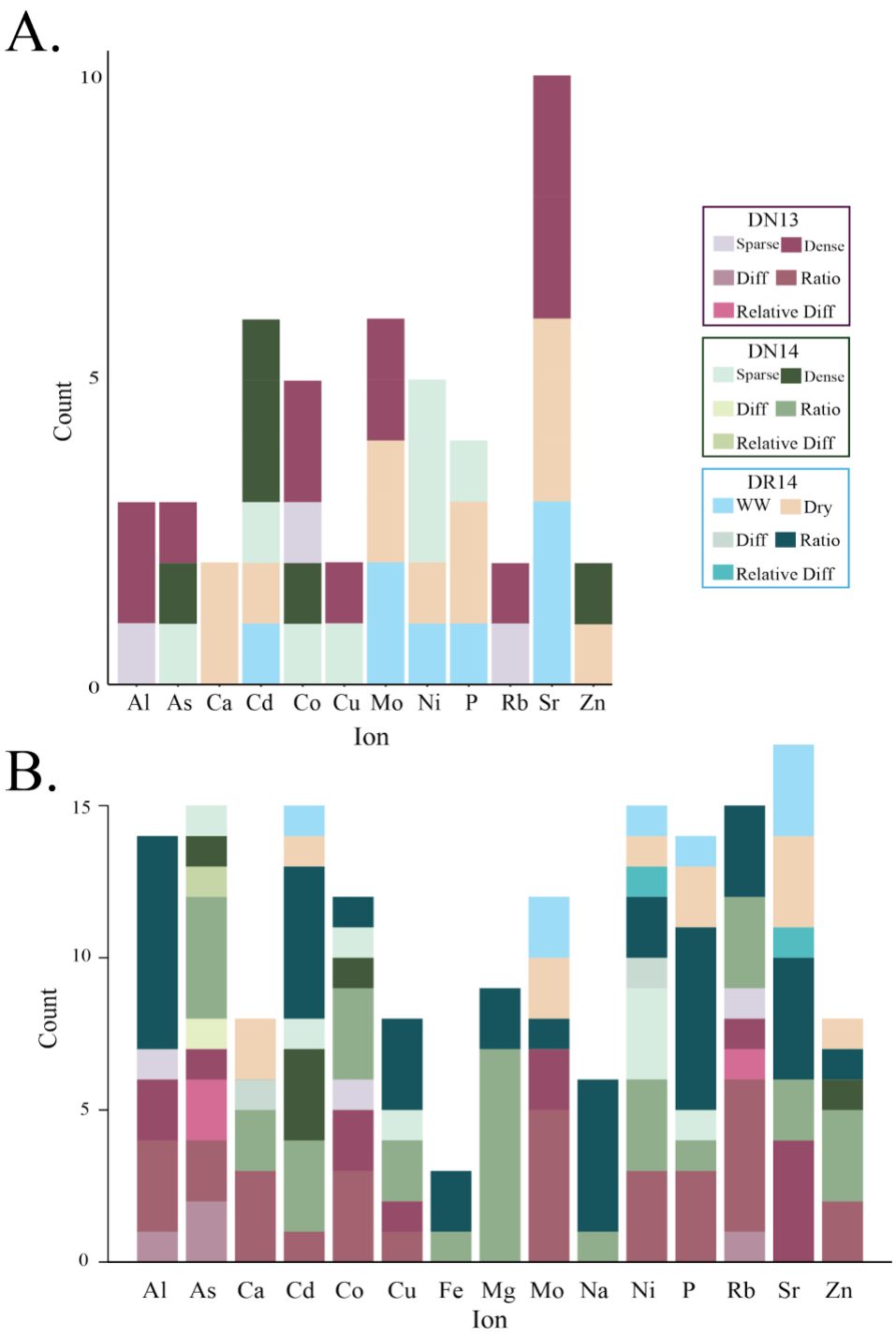


Figure 3-2. Count of ion specific QTL

Colors indicate the phenotypic metric for which the QTL were identified.

Figure 3-3

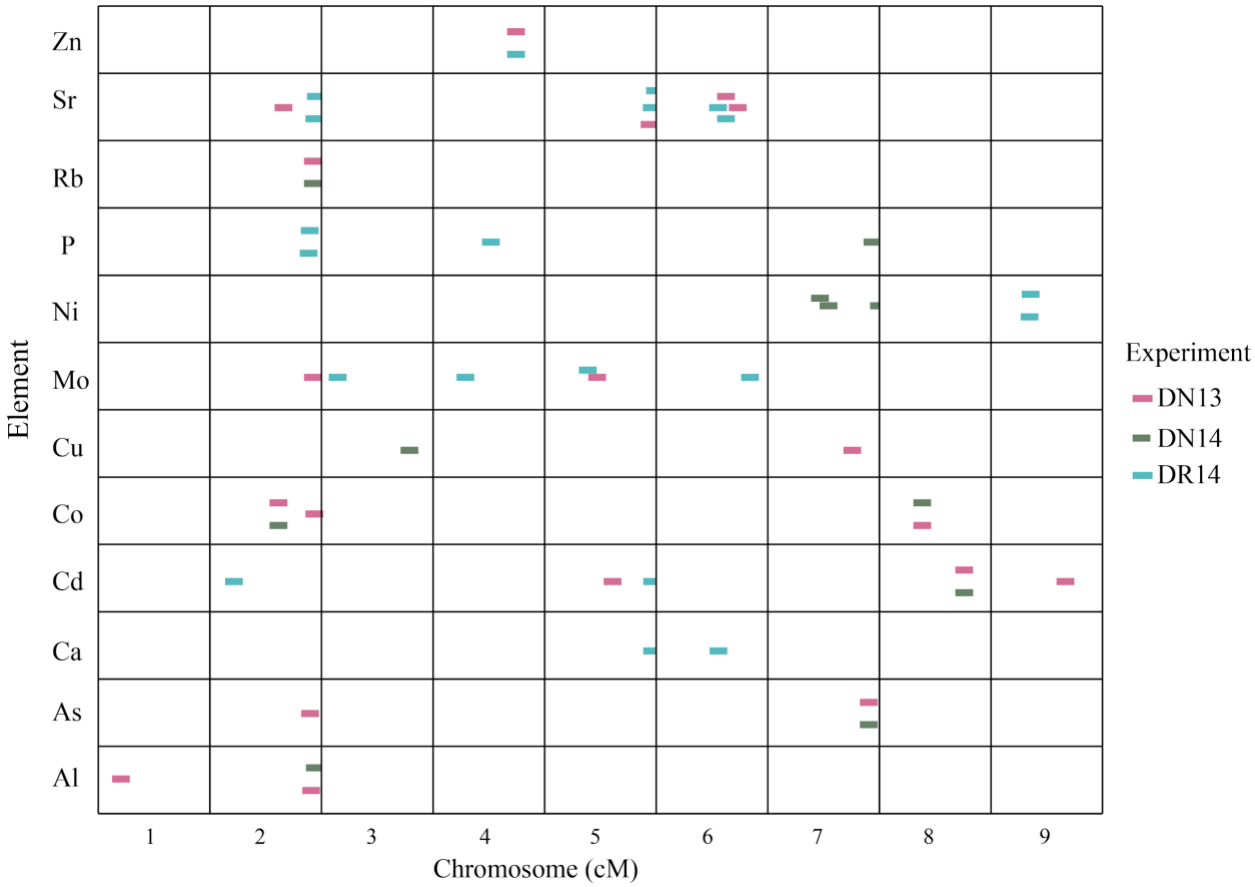


Figure 3-3. Genomic location of ion specific QTL mapped using traits from single treatments

Colors indicate the experiment in which the specific QTL was identified. QTL that were within 5 cM of each other, and were identified for the same phenotype (e.g. densely planted As content), were considered the same QTL. A QTL at the end of chromosome 2 was identified for seven ions.

Figure 3.4

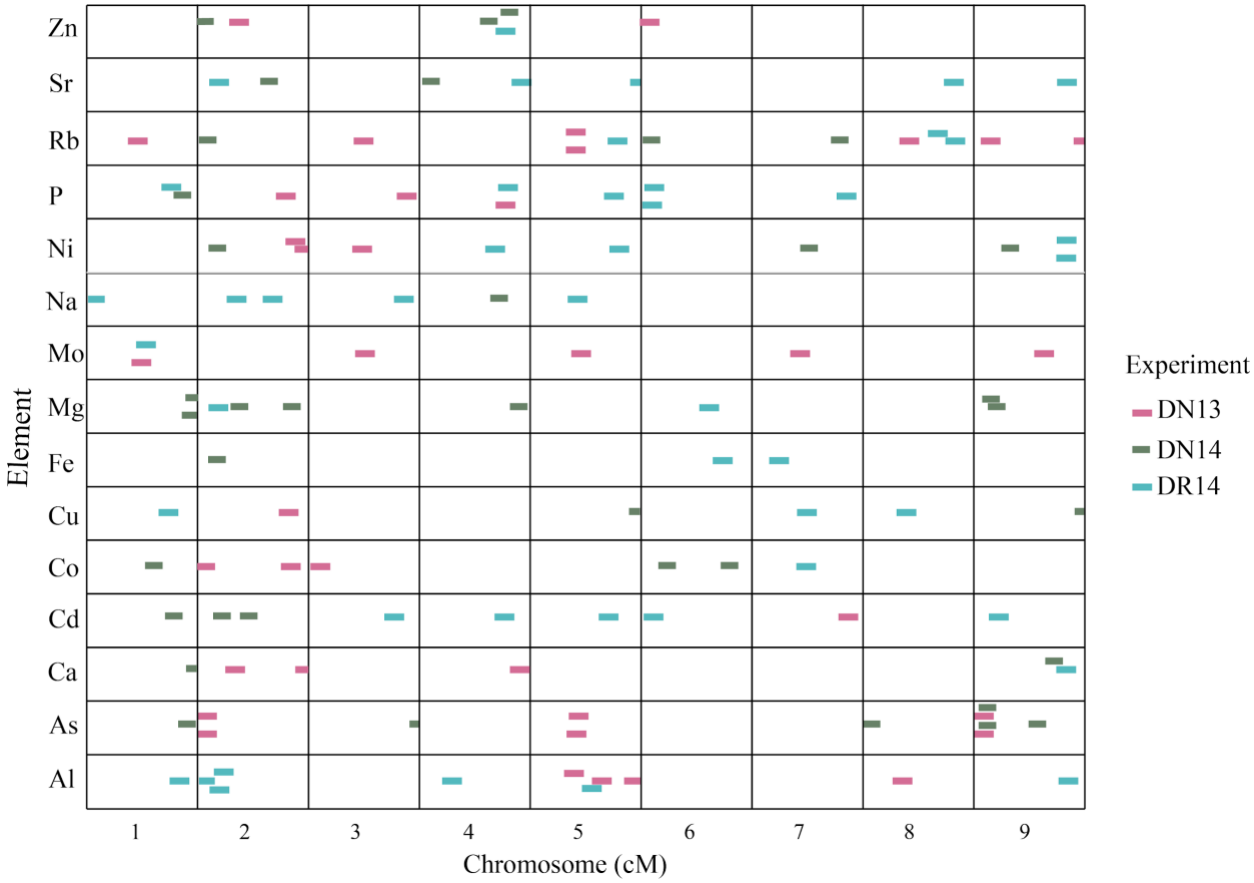


Figure 3.4. Genomic location of ion specific QTL mapped using differential traits

Figure 3-5

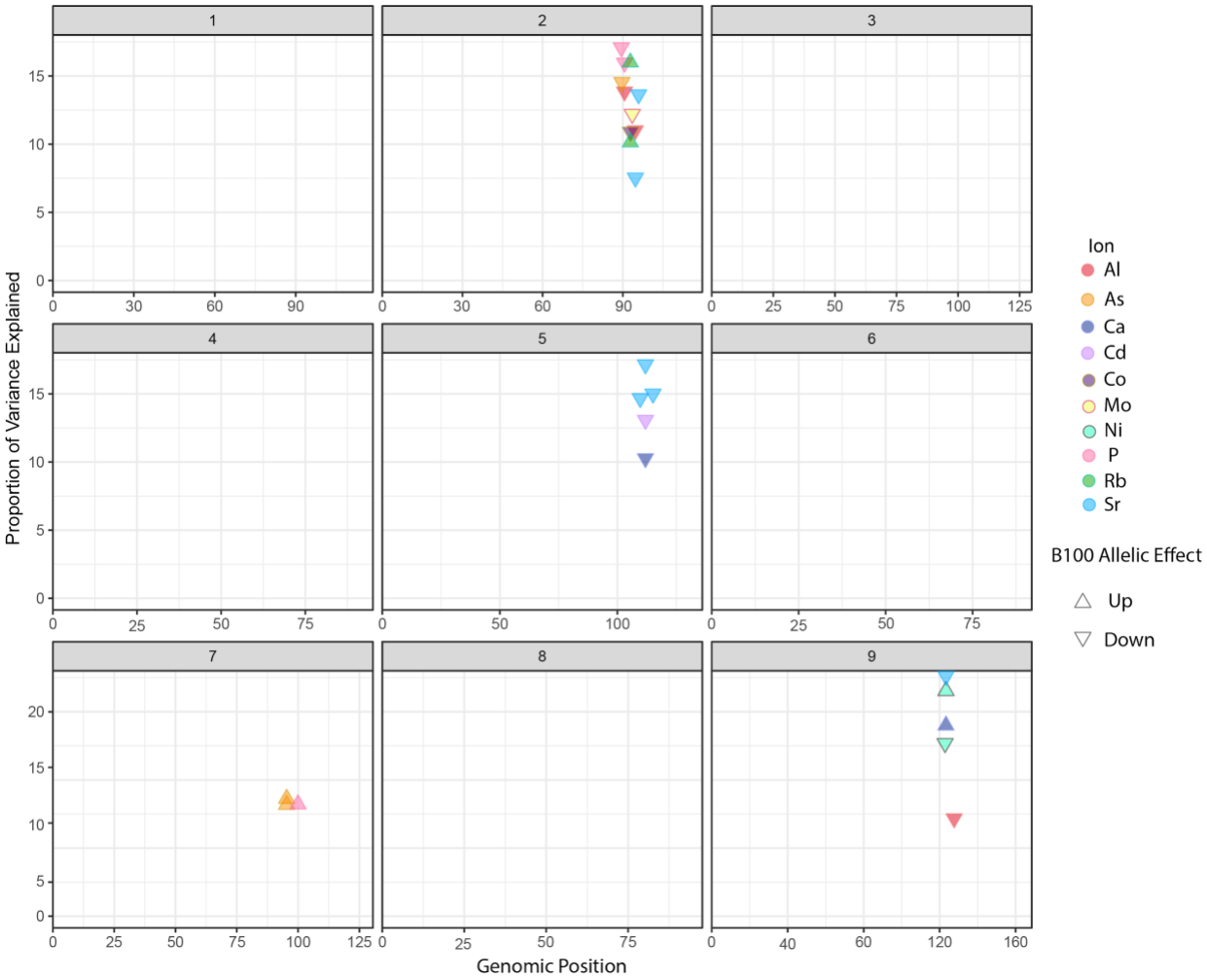


Figure 3-5. Stacks of QTL for multiple ions were found on chromosomes 2, 5, 7, and 9

The B100 haplotype at these loci was typically associated with a decrease in concentration for the element of interest. Notable exceptions to this trend are the relative increase of P and As due to the B100 haplotype on chromosome 7, as well as in Rb on chromosome 2.

Figure 3-6

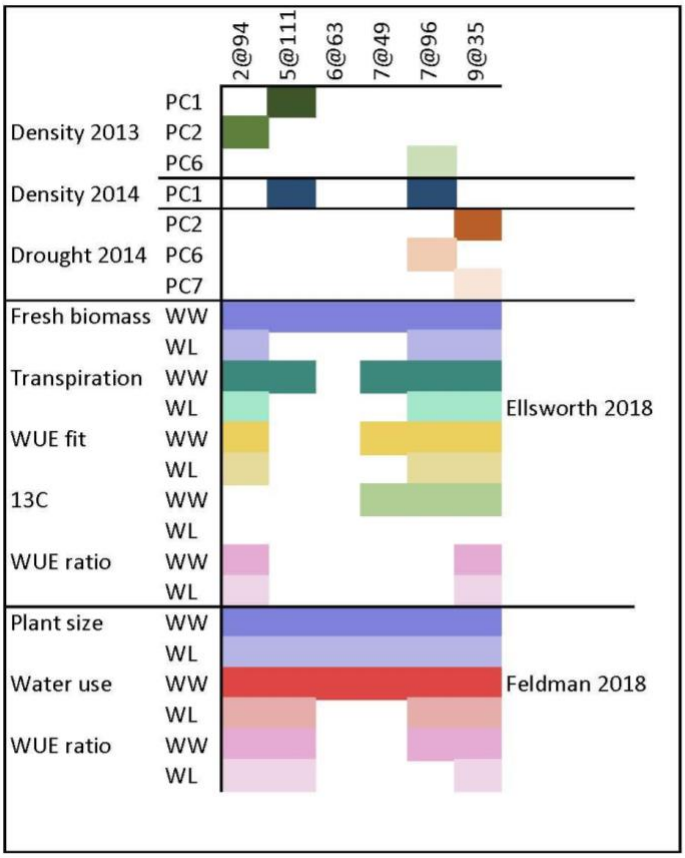


Figure 3-6. Ionomic QTL identified using principal components analysis overlap regions associated with water use efficiency

Traits assayed include WUE_{ratio} (the ratio of aboveground biomass and total water transpired), WUE_{fit} (accumulation of fresh biomass based on the amount of water transpired), and 13C (concentration of Carbon 13). Each trait was assessed in a well watered (WW) and water limited (WL) setting.

Table 3-1

Element	DN13	DN14	DR14	Repeatability
Al	0.4394728	0.357870473	0.581854268	0.300557557
As	0.2519856	0.327438597	0.570151195	0.248128951
B	0.1071448	0.225769155	0.166526725	0.093471981
Ca	0.3199173	0.325503874	0.45723887	0.237425136
Cd	0.442759	0.351500239	0.605383773	0.286845658
Co	0.4544401	0.343000169	0.442721132	0.230139108
Cu	0.2200043	0.367052581	0.723380607	0.43800547
Fe	0.0194658	0.326208207	0.474322266	0.265574993
K	0.4240101	0.364287029	0.713975561	0.272374312
Mg	0.0833627	0.351590025	0.50735422	0.178953196
Mn	0.399126	0.371214288	0.632785949	0.299522693
Mo	0.3766015	0.21450294	0.623190351	0.145177631
Na	0	0.255567401	0.184714412	0.152057979
Ni	0.2674639	0.223940266	0.415996303	0.20900581
P	0.1763122	0.33195116	0.707807003	0.350512411
Rb	0.441271	0.346945671	0.535080193	0.226641635
S	0.2903022	0.351714158	0.672582601	0
Se	0.1068793	0.053215905	0.182536896	0.132657003
Sr	0.2903022	0.351714158	0.672582601	0.29098477
Zn	0.3252552	0.39304282	0.472027959	0.300250792

Table 3-1. Heritability and repeatability for 20 ions

Table 3-2

Position	LOD	Trait	Treatment	Experiment Type	Year
1@20.5	3.6	Al	Dense	Density	2013
1@87.1	7.2	Al	Differential	Drought	2014
2@22.4	30.1	Al	Differential	Drought	2014
2@27.1	22.3	Al	Differential	Drought	2014
2@4	14.2	Al	Differential	Drought	2014
2@90.6	4.9	Al	Dense	Density	2013
2@94	3.3	Al	Sparse	Density	2013
4@29.5	27.8	Al	Differential	Drought	2014
5@40.6	3.2	Al	Differential	Density	2013
5@57.1	22.9	Al	Differential	Drought	2014
5@66.4	82.8	Al	Differential	Density	2013
5@96.2	8.6	Al	Differential	Density	2013
8@35.4	24.4	Al	Differential	Density	2013
9@126.5	13.8	Al	Differential	Drought	2014
1@95.2	7.3	As	Differential	Density	2014
2@3.2	55	As	Differential	Density	2013
2@3.7	53.2	As	Differential	Density	2013
2@89.7	4.6	As	Dense	Density	2013
3@120	6.3	As	Differential	Density	2014
5@42.8	4.2	As	Differential	Density	2013
5@44.7	4.2	As	Differential	Density	2013
7@95.3	4.7	As	Dense	Density	2014
7@95.3	4.5	As	Sparse	Density	2014
8@0.9	43.2	As	Differential	Density	2014
9@11.1	3.4	As	Differential	Density	2013
9@11.1	4.2	As	Differential	Density	2013
9@17.3	5.2	As	Differential	Density	2014
9@17.3	5.2	As	Differential	Density	2014

9@86	64.9	As	Differential	Density	2014
1@104.3	9.3	Ca	Differential	Density	2014
2@114.6	47.6	Ca	Differential	Density	2013
2@38.9	25.1	Ca	Differential	Density	2013
4@90.5	22.2	Ca	Differential	Density	2013
5@111.9	3.6	Ca	Drought	Drought	2014
6@56.5	5.7	Ca	Drought	Drought	2014
9@108.8	8.7	Ca	Differential	Density	2014
9@123.7	6.1	Ca	Differential	Drought	2014
1@82	49.6	Cd	Differential	Density	2014
2@22.4	4.3	Cd	Well Watered	Drought	2014
2@25.2	95.3	Cd	Differential	Density	2014
2@52.9	68.1	Cd	Differential	Density	2014
3@91.8	71.6	Cd	Differential	Drought	2014
4@75.9	4.1	Cd	Differential	Drought	2014
5@111.9	3.9	Cd	Drought	Drought	2014
5@71	4	Cd	Dense	Density	2014
5@71.9	3.1	Cd	Differential	Drought	2014
6@10.3	17.6	Cd	Differential	Drought	2014
7@85.8	9.2	Cd	Differential	Density	2013
8@76.8	4.2	Cd	Dense	Density	2014
8@76.8	4	Cd	Sparse	Density	2014
9@32	88.2	Cd	Differential	Drought	2014
9@67.6	4.2	Cd	Dense	Density	2014
1@63.9	9.5	Co	Differential	Density	2014
2@61.6	3.3	Co	Dense	Density	2014
2@61.6	3.2	Co	Sparse	Density	2014
2@7.3	46.1	Co	Differential	Density	2013
2@93.5	3.9	Co	Dense	Density	2013
2@95.9	32.9	Co	Differential	Density	2013

3@12.6	27.4	Co	Differential	Density	2013
6@24	7.9	Co	Differential	Density	2014
6@80.2	58.3	Co	Differential	Density	2014
7@48.6	27.4	Co	Differential	Drought	2014
8@38.6	3.3	Co	Dense	Density	2013
8@38.6	4.1	Co	Sparse	Density	2013
1@76.4	5.1	Cu	Differential	Drought	2014
2@93.5	10.7	Cu	Differential	Density	2013
3@79.8	4	Cu	Sparse	Density	2014
5@100.4	8.1	Cu	Differential	Density	2014
7@49.1	47.3	Cu	Differential	Drought	2014
7@80.6	3.6	Cu	Dense	Density	2013
8@38.6	26.4	Cu	Differential	Drought	2014
9@147.6	16.2	Cu	Differential	Density	2014
2@20.9	10.8	Fe	Differential	Density	2014
6@73.3	22.8	Fe	Differential	Drought	2014
7@24.4	68.1	Fe	Differential	Drought	2014
1@102.6	11.5	Mg	Differential	Density	2014
1@99	11.6	Mg	Differential	Density	2014
2@21.7	14	Mg	Differential	Drought	2014
2@44.4	18.4	Mg	Differential	Density	2014
2@98.5	26.4	Mg	Differential	Density	2014
4@90.5	20.1	Mg	Differential	Density	2014
6@61.4	8.9	Mg	Differential	Drought	2014
9@24.4	31.4	Mg	Differential	Density	2014
9@32	46.9	Mg	Differential	Density	2014
1@51.2	7.7	Mo	Differential	Density	2013
1@55.4	44.8	Mo	Differential	Drought	2014
2@93.5	4.5	Mo	Dense	Density	2013
3@15.5	3.3	Mo	Drought	Drought	2014

3@60.8	11.5	Mo	Differential	Density	2013
4@30	5.3	Mo	Drought	Drought	2014
5@45.9	3.9	Mo	Well Watered	Drought	2014
5@46.9	13.5	Mo	Differential	Density	2013
5@55.4	4.6	Mo	Dense	Density	2013
6@85.2	3.4	Mo	Well Watered	Drought	2014
7@43	25.8	Mo	Differential	Density	2013
9@93.7	80.9	Mo	Differential	Density	2013
1@8.4	12.5	Na	Differential	Drought	2014
2@41	6.5	Na	Differential	Drought	2014
2@78.2	8.4	Na	Differential	Drought	2014
3@104.3	236.9	Na	Differential	Drought	2014
4@73.7	315.7	Na	Differential	Density	2014
5@44.7	6.6	Na	Differential	Drought	2014
2@100.6	6.1	Ni	Differential	Density	2013
2@110.1	8.6	Ni	Differential	Density	2013
2@20.9	62.5	Ni	Differential	Density	2014
3@57.6	3.6	Ni	Differential	Density	2013
4@67.9	6.3	Ni	Differential	Drought	2014
5@82.1	13.2	Ni	Differential	Drought	2014
7@105.3	6.2	Ni	Sparse	Density	2014
7@49.8	5.2	Ni	Sparse	Density	2014
7@51.9	35	Ni	Differential	Density	2014
7@57.7	6.2	Ni	Sparse	Density	2014
9@123.2	5.5	Ni	Differential	Drought	2014
9@123.7	7.4	Ni	Differential	Drought	2014
9@34.9	8	Ni	Well Watered	Drought	2014
9@35.9	7.5	Ni	Drought	Drought	2014
9@49.7	16.2	Ni	Differential	Density	2014
1@78.3	16.7	P	Differential	Drought	2014

1@89.5	25.8	P	Differential	Density	2014
2@89.4	58.2	P	Differential	Density	2013
2@89.4	5.9	P	Drought	Drought	2014
2@90.6	5.2	P	Well Watered	Drought	2014
3@104.5	9.6	P	Differential	Density	2013
4@52.9	3.8	P	Drought	Drought	2014
4@76.2	74.6	P	Differential	Density	2013
4@78.5	24.2	P	Differential	Drought	2014
5@76.2	87.9	P	Differential	Drought	2014
6@10.3	11.6	P	Differential	Drought	2014
6@8.3	47.4	P	Differential	Drought	2014
7@83.6	106.3	P	Differential	Drought	2014
7@99.9	4.5	P	Sparse	Density	2014
1@46.3	29.5	Rb	Differential	Density	2013
2@4	56.4	Rb	Differential	Density	2014
2@92.8	5.1	Rb	Dense	Density	2013
2@92.8	3.2	Rb	Sparse	Density	2013
3@57.6	54	Rb	Differential	Density	2013
5@40.3	3.2	Rb	Differential	Density	2013
5@40.3	3.2	Rb	Differential	Density	2013
5@79.2	23.3	Rb	Differential	Drought	2014
6@0.7	52.1	Rb	Differential	Density	2014
7@78.1	226	Rb	Differential	Density	2014
8@40.1	28.3	Rb	Differential	Density	2013
8@65.6	19.8	Rb	Differential	Drought	2014
8@81.6	8.2	Rb	Differential	Drought	2014
9@148.5	116.5	Rb	Differential	Density	2013
9@19.9	77.3	Rb	Differential	Density	2013
2@21.7	53	Sr	Differential	Drought	2014
2@67	5.4	Sr	Dense	Density	2013

2@74.2	54.3	Sr	Differential	Density	2014
2@94.7	3.3	Sr	Drought	Drought	2014
2@95.9	5.6	Sr	Well Watered	Drought	2014
4@10.9	23.8	Sr	Differential	Density	2014
4@91.5	63.8	Sr	Differential	Drought	2014
5@102.9	41.9	Sr	Differential	Drought	2014
5@109.7	6	Sr	Dense	Density	2013
5@111.9	7.1	Sr	Drought	Drought	2014
5@115.1	6.1	Sr	Well Watered	Drought	2014
6@56.5	9	Sr	Drought	Drought	2014
6@63.7	7	Sr	Dense	Density	2013
6@63.7	5.5	Sr	Well Watered	Drought	2014
6@74.1	3.9	Sr	Dense	Density	2013
8@81.6	22.3	Sr	Differential	Drought	2014
9@123.7	7.9	Sr	Differential	Drought	2014
2@4	33.3	Zn	Differential	Density	2014
2@43.7	100.6	Zn	Differential	Density	2013
4@64.3	96.2	Zn	Differential	Density	2014
4@73.7	4.3	Zn	Drought	Drought	2014
4@74	3.2	Zn	Dense	Density	2014
4@78.5	23.1	Zn	Differential	Drought	2014
4@82.9	12.7	Zn	Differential	Density	2014
6@8.3	18.7	Zn	Differential	Density	2013
2@20.9	11.4	PC1	Differential	Density	2013
5@111.9	3.3	PC1	Sparse	Density	2013
5@55.4	3.8	PC1	Well Watered	Drought	2014
6@90.8	128.2	PC1	Differential	Density	2014
7@2.2	88.9	PC1	Differential	Density	2014
7@99.9	5.2	PC1	Sparse	Density	2014
2@39.9	5.7	PC10	Drought	Drought	2014

2@39.9	3.5	PC10	Well Watered	Drought	2014
2@58.9	12.9	PC10	Differential	Drought	2014
5@88.4	11.6	PC10	Differential	Drought	2014
3@89.5	62.8	PC11	Differential	Drought	2014
9@88.2	39.4	PC11	Differential	Drought	2014
1@75.9	3.6	PC2	Sparse	Density	2014
2@90.6	4.6	PC2	Dense	Density	2013
4@0.2	3.2	PC2	Dense	Density	2013
4@93.7	24.1	PC2	Differential	Drought	2014
5@39.4	3.3	PC2	Sparse	Density	2013
7@43	17.3	PC2	Differential	Drought	2014
8@32.5	5.3	PC2	Sparse	Density	2013
8@35.9	3.2	PC2	Dense	Density	2013
9@29.4	17.5	PC2	Differential	Drought	2014
4@11.2	19	PC3	Differential	Density	2014
5@26.9	4.5	PC3	Drought	Drought	2014
5@37.4	5	PC3	Well Watered	Drought	2014
8@29.2	5.2	PC3	Drought	Drought	2014
9@67.7	9.7	PC3	Differential	Density	2014
2@77.8	3.4	PC4	Dense	Density	2013
3@58.1	5.2	PC4	Drought	Drought	2014
4@49.8	13.5	PC4	Differential	Drought	2014
4@70.5	29.1	PC4	Differential	Density	2014
4@91.5	6.8	PC4	Differential	Drought	2014
7@23.7	8	PC4	Differential	Drought	2014
7@65.1	19.4	PC4	Differential	Density	2014
7@66.6	4	PC4	Sparse	Density	2014
1@99.7	55.3	PC5	Differential	Density	2013
2@119.8	3.7	PC5	Sparse	Density	2013
2@81.5	28	PC5	Differential	Density	2014

4@1.2	26.3	PC5	Differential	Density	2013
5@96.7	27.3	PC5	Differential	Drought	2014
5@97.2	102.5	PC5	Differential	Density	2013
8@85.9	19.2	PC5	Differential	Density	2013
9@136.5	9.1	PC5	Differential	Drought	2014
2@14.7	8.9	PC6	Differential	Drought	2014
7@41	92.2	PC6	Differential	Drought	2014
7@80.6	4.8	PC6	Drought	Drought	2014
7@94.9	7.9	PC6	Differential	Drought	2014
7@99.9	4.4	PC6	Dense	Density	2013
7@99.9	3.6	PC6	Sparse	Density	2013
9@116.2	20.1	PC6	Differential	Drought	2014
9@17.3	66.9	PC6	Differential	Drought	2014
2@81.3	4.7	PC7	Dense	Density	2014
2@81.3	4.3	PC7	Sparse	Density	2014
3@107.4	6.4	PC7	Dense	Density	2014
3@119	5.5	PC7	Dense	Density	2014
3@15.5	62.3	PC7	Differential	Density	2014
3@60.4	33.8	PC7	Differential	Density	2014
4@75.7	14.2	PC7	Differential	Drought	2014
4@76.2	78.6	PC7	Differential	Density	2014
5@21.5	76.8	PC7	Differential	Density	2014
5@47.1	3.2	PC7	Dense	Density	2013
5@59.4	23.5	PC7	Differential	Density	2013
7@22.7	11.9	PC7	Differential	Drought	2014
7@70.1	4.8	PC7	Dense	Density	2013
7@78.4	63.9	PC7	Differential	Density	2014
8@79.6	19.1	PC7	Differential	Density	2014
8@81.1	9.7	PC7	Differential	Density	2013
9@39.6	3.2	PC7	Drought	Drought	2014

1@46	12.7	PC8	Differential	Drought	2014
1@79.2	20.2	PC8	Differential	Density	2013
2@74.5	9.1	PC8	Differential	Drought	2014
5@36.4	22.9	PC8	Differential	Drought	2014
2@82.7	98.6	PC9	Differential	Drought	2014
3@48.1	25.1	PC9	Differential	Drought	2014
3@60.4	24.1	PC9	Differential	Drought	2014
3@83.6	7	PC9	Differential	Density	2013
3@93.2	3.7	PC9	Drought	Drought	2014
4@36.6	4.5	PC9	Drought	Drought	2014
4@73.7	4.3	PC9	Well Watered	Drought	2014
7@0	15.7	PC9	Differential	Drought	2014
9@123.7	13.2	PC9	Differential	Drought	2014

Table 3-2. Positions of QTL identified for the ionome in *Setaria*

Table 3-3

QTL	Length	Observed Value	Additive Lower 95% CI	Additive Upper 95% CI	Trait	Known ion homeostasis genes
5@42.8	5343667	1	0.12	0.25	As DN13	Sevir.5G106900
9@11.1	354121	0				
9@11.1	354121	0	0.36	0.94	As DN13	Sevir.5G106900, Sevir.5G144600
5@44.7	28967705	2				
8@38.6	27998232	0	0.52	1.07	Co DN13	Sevir.2G352900, Sevir.2G374000
2@93.5	4061238	2				
7@80.6	6600579	1	0.08	0.22	Cu DN13	Sevir.7G206000
1@51.2	1612753	0	0.11	0.18	Mo DN13	Sevir.5G106900
3@60.8	1212521	0				
5@46.9	1872544	1				
7@43	507448	0				
9@93.7	477509	0				
2@93.5	883410	0	0.21	0.62	Mo DN13	Sevir.5G106900
5@55.4	25239042	1				
4@74	1166386	1	0.1	0.16	Zn DN14	Sevir.4G265200
6@85.2	1386783	0	0.26	0.73	Mo DR14	Sevir.5G106900, Sevir.5G251200

5@45.9	29040921	1				
4@52.9	21056782	1	0.83	1.4	P DR14	Sevir.4G173800
2@89.4	1722188	0				
4@73.7	711696	1	0.06	0.1	Zn DR14	Sevir.4G265200
5@111.9	39790077	14	8.92	13.44	PC1 DN13	Sevir.5G086900, Sevir.5G106900, Sevir.5G195200, Sevir.5G196600, Sevir.5G252300, Sevir.5G210300, Sevir.5G279700, Sevir.5G379500, Sevir.5G395800, Sevir.5G400300, Sevir.5G404100, Sevir.5G434600, Sevir.5G144600, Sevir.5G337200
4@0.2	8106883	1	3.25	3.91	PC2 DN13	Sevir.8G071400
8@35.9	825662	0				
2@90.6	1444572	0				
5@39.4	39222525	13	10.46	14.7	PC2 DN13	Sevir.5G086900, Sevir.5G106900, Sevir.5G195200, Sevir.5G196600, Sevir.5G252300, Sevir.5G210300, Sevir.5G279700, Sevir.5G379500, Sevir.5G395800, Sevir.5G400300, Sevir.5G404100, Sevir.5G144600, Sevir.5G337400
8@32.5	4776098	0				
2@77.8	4763267	3	1.5	1.89	PC4 DN13	Sevir.2G338800, Sevir.2G352900, Sevir.2G374000
2@119.8	910010	1	0.3	0.4	PC5 DN13	Sevir.2G452700
8@85.9	533540	0	0.56	0.66	PC5 DN13	Sevir.1G349900, Sevir.5G395800
4@1.2	218696	0				
1@99.7	371042	1				
5@97.2	442167	1				

8@81.1	1093914	0	1.04	1.25	PC7 DN13	Sevir.5G195200, Sevir.5G196600
5@59.4	1915491	2				
5@47.1	18215279	4	5.28	6.4	PC7 DN13	Sevir.5G106900, Sevir.5G195200, Sevir.5G196600, Sevir.5G144700, Sevir.7G205900
7@70.1	3016505	1				
1@79.2	469655	1	0.16	0.21	PC8 DN13	Sevir.1G292400
1@75.9	3547293	3	1.14	1.44	PC2 DN14	Sevir.1G275100, Sevir.1G292400, Sevir.1G301800
9@67.7	1127483	1	0.49	0.61	PC3 DN14	Sevir.9G238300
4@11.2	301729	0				
7@65.1	1401883	0	0.77	0.92	PC4 DN14	Sevir.4G249800, Sevir.4G255000
4@70.5	798551	2				
7@66.6	13176556	4	3.44	4.37	PC4 DN14	Sevir.7G134800, Sevir.7G245900, Sevir.7G252600, Sevir.7G195200
2@81.5	356827	1	0.12	0.16	PC5 DN14	Sevir.2G352900
8@79.6	1017166	0	1.26	1.42	PC7 DN14	Sevir.7G252600, Sevir.4G265200
3@60.4	528457	0				
3@15.5	335244	0				
7@78.4	438419	1				
5@21.5	333415	0				
4@76.2	829321	1				
2@81.3	4202757	3	1.78	2.15	PC7 DN14	Sevir.2G338800, Sevir.2G352900, Sevir.2G374000

3@119	945963	0				
3@107.4	230095	0				
2@81.3	14086974	4	3.59	4.53	PC7 DN14	Sevir.2G338800, Sevir.2G352900, Sevir.2G374000, Sevir.2G412100
5@55.4	19794116	5	4.43	5.59	PC1 DR14	Sevir.5G086900, Sevir.5G106900, Sevir.5G195200, Sevir.5G196600, Sevir.5G144600
5@37.4	39503672	13	8.96	13.15	PC3 DR14	Sevir.5G086900, Sevir.5G106900, Sevir.5G195200, Sevir.5G196600, Sevir.5G252300, Sevir.5G210300, Sevir.5G279700, Sevir.5G379500, Sevir.5G395800, Sevir.5G400300, Sevir.5G404100, Sevir.5G144600, Sevir.5G337400
5@26.9	5672938	1	2.81	3.32	PC3 DR14	Sevir.5G086900
8@29.2	2911543	0				
3@58.1	9731371	5	2.78	3.51	PC4 DR14	Sevir.3G122500, Sevir.3G153100, Sevir.3G159900, Sevir.3G213300, Sevir.3G220300
4@91.5	89563	1	0.93	1.15	PC4 DR14	Sevir.4G295200
7@23.7	199726	0				
4@49.8	2493109	0				
7@80.6	2263241	1	0.75	0.96	PC6 DR14	Sevir.7G252600
7@94.9	373524	0	7.53	9.46	PC6 DR14	Sevir.2G037000, Sevir.2G125000, Sevir.2G205200, Sevir.2G228000, Sevir.2G071500, Sevir.2G092800, Sevir.9G044800, Sevir.7G089700
2@14.7	30124161	6				
9@116.2	771696	0				
9@17.3	829584	1				

7@41	1034133	1				
9@39.6	2637311	3	0.86	1.1	PC7 DR14	Sevir.9G133800, Sevir.9G133900, Sevir.9G144800
7@22.7	646682	0	0.48	0.59	PC7 DR14	Sevir.4G265200
4@75.7	730479	1				
4@73.7	987170	2	0.33	0.43	PC9 DR14	Sevir.4G255000, Sevir.4G265200
3@93.3	1878840	1	8.09	10.6	PC9 DR14	Sevir.4G173800, Sevir.4G136300, Sevir.4G150000, Sevir.4G231900, Sevir.4G249800, Sevir.4G255000, Sevir.4G039600, Sevir.3G384700
4@36.6	33757934	7				
9@123.7	541959	0	0.87	1.01	PC9 DR14	Sevir.3G122500
7@0	605215	0				
3@48.1	644103	1				
2@82.7	654940	0				

Table 3-3. SPQV results and ionomic genes identified by QTL

Supplemental Table 3-1

	Method		
Trait	DN13	DN14	DR14
Al	max likelihood lambda	lognormal	max likelihood lambda
As	lognormal	lognormal	max likelihood lambda
Br	sqrt transformation	max likelihood lambda	max likelihood lambda
Ca	lognormal	lognormal	sqrt transformation
Cd	lognormal	max likelihood lambda	max likelihood lambda
Co	max likelihood lambda	lognormal	lognormal
Cu	lognormal	normal	lognormal
Fe	lognormal	lognormal	lognormal
K	sqrt transformation	max likelihood lambda	lognormal
Mg	lognormal	max likelihood lambda	lognormal
Mn	lognormal	max likelihood lambda	lognormal
Mo	lognormal	lognormal	lognormal
Na	sqrt transformation	max likelihood lambda	max likelihood lambda
Ni	lognormal	lognormal	lognormal
P	Inverse sqrt transformation	sqrt transformation	lognormal
Rb	max likelihood lambda	lognormal	max likelihood lambda
S	sqrt transformation	max likelihood lambda	max likelihood lambda
Se	normal	normal	max likelihood lambda
Sr	lognormal	lognormal	max likelihood lambda
Zn	sqrt transformation	lognormal	lognormal

Supplemental Table 3-1. Transformations applied to ionomic data to achieve normality

Supplemental Table 3-2

GeneID	Elements	Gene inferred from	A.thaliana orthologs
Sevir.1G132950	Se, S	AT1G08490	AT1G08490
Sevir.1G275100	P	AT1G12640	AT1G12640, AT1G63050
Sevir.1G212900	Zn	AT1G14870	AT1G14870, AT1G14880, AT5G35525
Sevir.1G213100	Zn	AT1G14870	AT1G14870, AT1G14880, AT5G35525
Sevir.2G374000	Fe, Zn, Co, Mn	AT1G20110	AT1G20110
Sevir.4G295200	Fe, Zn, Co, Mn	AT1G20110	AT1G20110
Sevir.2G037000	K	AT1G30270	AT1G30270
Sevir.7G252600	Cd, As	AT1G30400, LOC_Os04g52900	AT1G30400, AT2G34660
Sevir.3G283200	Ca, K, Na, S	AT1G30450	AT1G30450
Sevir.5G024400	Ca, K, Na, S	AT1G30450	AT1G30450
Sevir.1G301800	K, NO3-	AT1G32450	AT1G32450, AT4G21680
Sevir.5G404100	S, Se	AT1G36370	AT1G22020, AT1G36370
Sevir.2G452700	Fe, Cd, Co, Mo	AT1G56430	AT1G09240, AT1G56430, AT5G04950, AT5G56080
Sevir.9G044800	Fe, Cd, Co, Mo	AT1G56430, LOC_Os03g19420	AT1G09240, AT1G56430, AT5G04950, AT5G56080
Sevir.4G150000	Cd	AT1G59870	AT1G15210, AT1G59870
Sevir.4G249800	Fe	AT1G60960	AT1G10970, AT1G60960
Sevir.2G338800	S, Se	AT1G62180	AT1G62180, AT4G04610, AT4G21990
Sevir.7G205900	Cu	AT1G63440, LOC_Os04g46940	AT1G63440
Sevir.7G206000	Cu	AT1G63440	AT1G63440
Sevir.8G115200	Cu	AT1G66240, LOC_Os08g10480	AT1G66240
Sevir.3G052300	P	AT1G68320	AT1G25340, AT1G68320
Sevir.3G348200	P	AT1G68320	AT1G25340, AT1G68320
Sevir.8G248300	P	AT1G68320	AT1G25340, AT1G68320
Sevir.9G429300	P	AT1G68320	AT1G25340, AT1G68320
Sevir.4G173800	P, As	AT1G76430	AT1G20860, AT1G76430
Sevir.4G276100	Mn	AT1G80830	AT1G15960, AT1G80830

Sevir.2G205200	Fe, Zn	AT2G01770, LOC_Os09g23300	AT2G01770
Sevir.7G134800	Fe, Zn	AT2G01770, LOC_Os04g38940	AT2G01770
Sevir.3G427000	Na	AT2G01980	AT2G01980
Sevir.9G409800	S	AT2G13540	AT2G13540
Sevir.4G293800	Zn	AT2G16770, AT4G35040	AT2G16770, AT4G35040
Sevir.9G058800	Zn	AT2G16770, AT4G35040	AT2G16770, AT4G35040
Sevir.4G255000	Zn	AT2G19110, AT4G30110	AT2G19110, AT4G30110
Sevir.3G384700	Fe, Mn, Zn	AT2G23150	AT1G47240, AT2G23150
Sevir.9G169700	Fe, Mn, Zn	AT2G23150	AT1G47240, AT2G23150
Sevir.9G343900	Cu, Zn	AT2G23240, AT2G42000	AT2G23240, AT2G42000
Sevir.9G183400	Mo	AT2G25680, Medtr1g010270, Medtr3g464210, LOC_Os08g01120	AT2G25680
Sevir.7G076300	Fe	AT2G28160	AT2G28160
Sevir.7G076400	Fe	AT2G28160	AT2G28160
Sevir.5G106900	Ca, Mn, Zn, Na, S, K, As, Se, Mo	AT2G28670	AT1G07730, AT2G28670
Sevir.7G019300	P	AT2G32830	AT2G32830, AT5G43360
Sevir.3G153100	P	AT2G33770, LOC_Os05g48390	AT2G33770
Sevir.4G281800	Co, Ni	AT2G38460, AT5G03570	AT2G38460, AT5G03570
Sevir.9G238300	P	AT2G38940	AT2G38940
Sevir.9G238400	P	AT2G38940	AT2G38940
Sevir.9G542900	P	AT2G38940	AT2G38940
Sevir.3G220300	Mn	AT2G39450	AT2G39450
Sevir.5G379500	Mn	AT2G39450	AT2G39450
Sevir.3G050000	Zn	AT2G46800	AT2G46800
Sevir.3G347400	B	AT2G47160, AT3G62270, LOC_Os12g37840	AT2G47160, AT3G62270
Sevir.9G133900	P	AT3G01310, AT5G15070	AT3G01310, AT5G15070
Sevir.1G292400	Na, K, Rb, Mg, Ca, Fe, Mo	AT3G06060	AT3G06060, AT5G19200
Sevir.6G025100	Mn	AT3G12750	AT3G12750
Sevir.7G245900	Mn, Zn	AT3G12750, LOC_Os04g52310	AT3G12750

Sevir.3G034900	Mn, Fe, K, P	AT3G13320	AT1G55720, AT1G55730, AT3G13320
Sevir.9G379800	Mn, Fe, K, P	AT3G13320	AT1G55720, AT1G55730, AT3G13320
Sevir.1G349900	S	AT3G14280	AT3G14280
Sevir.4G039500	S	AT3G14280	AT3G14280
Sevir.4G039600	S	AT3G14280	AT3G14280
Sevir.3G014100	Na, Fe, Zn, Mn, Mo	AT3G15380	AT3G15380
Sevir.6G001500	Na, Fe, Zn, Mn, Mo	AT3G15380	AT3G15380
Sevir.3G159900	Fe, Zn, Mn	AT3G18290	AT3G18290
Sevir.5G279700	Fe, Zn, Mn	AT3G18290	AT3G18290
Sevir.9G095300	S	AT3G22890	AT3G22890, AT4G14680, AT5G43780
Sevir.1G367100	P	AT3G23430, LOC_Os02g56510	AT3G23430
Sevir.2G412100	Fe, Zn, Mn, Co	AT3G47640	AT3G47640
Sevir.8G189400	Fe, Zn, Mn, Co	AT3G47640	AT3G47640
Sevir.9G389000	Fe, Zn, Mn, Co	AT3G47640	AT3G47640
Sevir.2G071500	Na	AT3G47950	AT3G47950, AT5G62670
Sevir.3G424800	Na	AT3G47950	AT3G47950, AT5G62670
Sevir.9G133800	Na	AT3G47950	AT3G47950, AT5G62670
Sevir.3G122500	P, K	AT3G51860	AT2G38170, AT3G51860
Sevir.5G196600	P, K	AT3G51860	AT2G38170, AT3G51860
Sevir.9G537000	S	AT3G51895	AT3G51895
Sevir.5G462400	Fe	AT3G56970, AT3G56980	AT2G41240, AT3G56970, AT3G56980, AT5G04150
Sevir.9G473100	Fe	AT3G56970, AT3G56980	AT2G41240, AT3G56970, AT3G56980, AT5G04150
Sevir.1G343700	Mn	AT3G58060, LOC_Os02g53490	AT3G58060
Sevir.9G482200	Mn	AT3G58060, LOC_Os03g12530	AT3G58060
Sevir.9G341300	Mg	AT3G58970	AT3G58970
Sevir.9G265300	Fe	AT4G02780	AT4G02780
Sevir.9G357500	Fe	AT4G02780	AT4G02780
Sevir.1G067600	Na	AT4G10310	AT4G10310

Sevir.9G193400	B	AT4G10380,GRMZM2G176209	AT4G10380
Sevir.5G446100	Rb, Cs, K	AT4G13420,GRMZM2G084779	AT4G13420
Sevir.9G088600	Fe, Cd	AT4G16370	AT4G16370
Sevir.9G144700	Fe, Mn, Co, Cd, Zn	AT4G19690	AT1G31260, AT4G19690
Sevir.9G144800	Fe, Mn, Co, Cd, Zn	AT4G19690	AT1G31260, AT4G19690
Sevir.3G050100	Cd, As	AT4G23100	AT4G23100
Sevir.2G125000	P	AT4G28610	AT2G20400, AT4G28610
Sevir.9G420400	P	AT4G28610, LOC_Os03g21240	AT2G20400, AT4G28610
Sevir.5G210300	K	AT4G33000	AT4G33000
Sevir.4G265200	Zn	AT4G37270	AT4G37270
Sevir.6G149500	As, P	AT5G03455	AT5G03455
Sevir.9G340600	As, P	AT5G03455	AT5G03455
Sevir.9G577800	As, P	AT5G03455	AT5G03455
Sevir.3G091300	Zn, Fe	AT5G13740	AT5G13740, AT5G13750
Sevir.7G322400	Zn, Fe	AT5G13740, LOC_Os12g03899	AT5G13740, AT5G13750
Sevir.8G023200	Zn, Fe	AT5G13740	AT5G13740, AT5G13750
Sevir.8G023400	Zn, Fe	AT5G13740, LOC_Os12g03899	AT5G13740, AT5G13750
Sevir.8G071400	Zn, Fe	AT5G13740	AT5G13740, AT5G13750
Sevir.9G083800	Ca, Mg	AT5G15410	AT5G15410
Sevir.1G112000	Fe, Mn, Zn	AT5G17290	AT5G17290
Sevir.3G251900	Cu	AT5G18830	AT5G18830
Sevir.2G228000	Cu	AT5G20650	AT5G20650
Sevir.4G231900	Na	AT5G35410	AT5G35410
Sevir.5G470600	Fe	AT5G42130	AT5G42130
Sevir.5G144600	Zn, Cd, As	AT5G44070	AT1G03980, AT5G44070
Sevir.5G144700	Zn, Cd, As	AT5G44070, LOC_Os06g01260	AT1G03980, AT5G44070
Sevir.1G014200	Pb	AT5G53130	AT5G53130
Sevir.4G136300	Pb	AT5G53130	AT5G53130
Sevir.7G195100	Fe, Zn, Cu	AT5G53550, LOC_Os04g45860	AT5G24380, AT5G53550
Sevir.7G195200	Fe, Zn, Cu	AT5G53550, LOC_Os04g45900	AT5G24380, AT5G53550

Sevir.2G352900	Cd, Co, Fe, Mn, Zn	AT5G54680	AT1G51070, AT5G54680
Sevir.6G068300	Cd, Co, Fe, Mn, Zn	AT5G54680, LOC_Os08g04390	AT1G51070, AT5G54680
Sevir.6G072000	Cd	AT5G54810	AT4G27070, AT5G54810
Sevir.6G105000	Li, B, Na, Mg, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, Rb, Sr, Mo, Cd	AT5G57620	AT5G57620
Sevir.5G337200	Cu	AT5G59030	AT2G26975, AT3G46900, AT5G59030, AT5G59040
Sevir.5G337400	Cu	AT5G59030	AT2G26975, AT3G46900, AT5G59030, AT5G59040
Sevir.1G340900	K	AT5G64930	AT5G64930
Sevir.5G434600	K	AT5G64930	AT5G64930
Sevir.2G103000	Fe, Mn, Cd	Medtr3g088460, LOC_Os07g15370	
Sevir.5G086900	Mn	LOC_Os01g03914	AT1G16310, AT1G79520
Sevir.5G023200	Na	LOC_Os01g20160,GRMZM2G047616	
Sevir.5G195200	Na	LOC_Os01g20160,GRMZM2G047616	
Sevir.5G252300	K	LOC_Os01g45990	AT2G26650
Sevir.5G395800	Fe	LOC_Os01g64250	AT3G54290
Sevir.5G400300	Mg, Na	LOC_Os01g64890	AT3G19640
Sevir.1G080600	As	LOC_Os02g06290	
Sevir.1G048900	Cu	LOC_Os02g10290	
Sevir.1G025500	As	LOC_Os02g13870	AT4G18910, AT4G19030
Sevir.1G258400	Fe, Mn	LOC_Os02g43370	
Sevir.1G258600	Fe	LOC_Os02g43410	
Sevir.1G325000	Se	LOC_Os02g51110	
Sevir.9G542800	Se	LOC_Os03g05640	
Sevir.9G514000	Fe, Zn	LOC_Os03g09140	
Sevir.9G441600	Fe	LOC_Os03g18550	AT1G07030, AT2G30160
Sevir.7G089700	Cs	LOC_Os04g32920	
Sevir.7G090200	Cs	LOC_Os04g32920	

Sevir.3G213300	Zn	LOC_Os05g39560	AT1G05300, AT2G32270
Sevir.4G031300	P	LOC_Os06g05160	AT3G15990
Sevir.8G124600	K	LOC_Os07g01810	AT5G55630
Sevir.2G068500	P	LOC_Os07g09000	AT3G52190
Sevir.2G092800	Cd	LOC_Os07g12900	
Sevir.2G093100	Cd	LOC_Os07g12900	
Sevir.6G060000	As	LOC_Os08g05590	
Sevir.9G479900	Fe	GRMZM2G060952	AT1G59950, AT1G59960

Supplemental Table 3-2. Known ionomic genes in *S. viridis*

Works Cited

Albretsen, J. (2006). The toxicity of iron, an essential element. *Veterinary Medicine - Bonner Springs,Edwardsville*. 101(2):82-90

Antonovics, J., Bradshaw, A. D., & Turner, R. G. (1971). Heavy Metal Tolerance in Plants. In J. B. Cragg (Ed.), *Advances in Ecological Research* (Vol. 7, pp. 1–85). Academic Press. https://doi.org/10.1016/S0065-2504(08)60202-0

Arends, D., Prins, P., Broman, K. W., & Jansen, R. C. (2016). Tutorial—Multiple-QTL Mapping (MQM) Analysis for R/qtl.

Assefa, T., Mahama, A. A., Brown, A. V., Cannon, E. K., Rubyogo, J. C., Rao, I. M., Blair, M. W., & Cannon, S. B. (2019). A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (Phaseolus vulgaris L.). *Molecular Breeding*, *39*(2), 20.

Bar-Ness, E., Hadar, Y., Chen, Y., Shanzer, A., & Libman, J. (1992). Iron uptake by plants from microbial siderophores: A study with 7-nitrobenz-2 oxa-1, 3-diazole-desferrioxamine as fluorescent ferrioxamine B analog. *Plant Physiology*, *99*(4), 1329–1335.

Barton, L., Newsome, S. D., Chen, F.-H., Wang, H., Guilderson, T. P., & Bettinger, R. L. (2009). Agricultural origins and the isotopic identity of domestication in northern China. *Proceedings of the National Academy of Sciences*, *106*(14), 5523–5528. https://doi.org/10.1073/pnas.0809960106

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baxter, I. R., Vitek, O., Lahner, B., Muthukumar, B., Borghi, M., Morrissey, J., Guerinot, M. L., & Salt, D. E. (2008). The leaf ionome as a multivariable system to detect a plant's physiological status. *Proceedings of the National Academy of Sciences*, *105*(33), 12081–12086. https://doi.org/10.1073/pnas.0804175105

Bell, G. (2012). *Selection: The Mechanism of Evolution*. Springer Science & Business Media.

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., Estep, M., Feng, L., Vaughn, J. N., Grimwood, J., Jenkins, J., Barry, K., Lindquist, E., Hellsten, U., Deshpande, S., Wang, X., Wu, X., Mitros, T., Triplett, J., … Devos, K. M. (2012). Reference genome sequence of the model plant *Setaria*. *Nature Biotechnology*, *30*(6), 555–561. https://doi.org/10.1038/nbt.2196

Bergen, B. K., & Chan Lau, T. T. (2012). Writing Direction Affects How People Map Space Onto Time. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00109

Bollard, E. G., & Butler, G. W. (1966). Mineral Nutrition of Plants. *Annual Review of Plant Physiology*, *17*(1), 77–112. https://doi.org/10.1146/annurev.pp.17.060166.000453

Boroditsky, L. (2001). Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, *43*(1), 1–22. https://doi.org/10.1006/cogp.2001.0748

Brady, N. C., & Weil, R. R. (2016). *The nature and properties of soils* (Vol. 13). Prentice Hall Upper Saddle River, NJ.

Brink, M., 2006. *Setaria italica* (L.) P. Beauv. Record from Protabase. Brink, M. & Belay, G. (Editors). PROTA (Plant Resources of Tropical Africa / Ressources végétales de l'Afrique tropicale), Wageningen, Netherlands

Broman, K. W. (2001). Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal*, *30*(7), 44–52.

Broman, Karl W., Wu, H., Sen, S., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics (Oxford, England)*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Brown, J. C., & Jones, W. E. (1975). Heavy-metal toxicity in plants. A crisis in embryo. *Communications in Soil Science and Plant Analysis*, *6*(4), 421–438. https://doi.org/10.1080/00103627509366579

Brutnell, T. P., Wang, L., Swartwood, K., Goldschmidt, A., Jackson, D., Zhu, X.-G., Kellogg, E., & Eck, J. V. (2010). *Setaria viridis*: A Model for C4 Photosynthesis. *The Plant Cell*, *22*(8), 2537–2544. https://doi.org/10.1105/tpc.110.075309

C. Davison, A., & Hinkley, D. (1997). *Bootstrap Methods and Their Application* (Vol. 94). https://doi.org/10.2307/1271471

Camp, S. D., Jolley, V. D., & Brown, J. C. (1987). Comparative evaluation of factors involved in Fe stress response in tomato and soybean. *Journal of Plant Nutrition*, *10*(4), 423–442. https://doi.org/10.1080/01904168709363583

Carey, A.-M., Norton, G. J., Deacon, C., Scheckel, K. G., Lombi, E., Punshon, T., Guerinot, M. L., Lanzirotti, A., Newville, M., Choi, Y., Price, A. H., & Meharg, A. A. (2011). Phloem transport of arsenic species from flag leaf to grain during grain filling. *New Phytologist*, *192*(1), 87–98. https://doi.org/10.1111/j.1469-8137.2011.03789.x

Casa, A. M., Pressoir, G., Brown, P. J., Mitchell, S. E., Rooney, W. L., Tuinstra, M. R., Franks, C. D., & Kresovich, S. (2008). Community resources and strategies for association mapping in sorghum. *Crop Science*, *48*(1), 30–40.

Ceasar, S. A., Hodge, A., Baker, A., & Baldwin, S. A. (2014). Phosphate Concentration and Arbuscular Mycorrhizal Colonisation Influence the Growth, Yield and Expression of Twelve PHT1 Family Phosphate Transporters in Foxtail Millet (*Setaria italica*). *PLOS ONE*, *9*(9), e108459. https://doi.org/10.1371/journal.pone.0108459

Celik, H., Asik, B. B., Gürel, S., & Katkat, A. V. (2010). Effects of potassium and iron on macro element uptake of maize. *Zemdirbyste 97(1)*

Chaney, R. L., Brown, J. C., & Tiffin, L. O. (1972). Obligatory reduction of ferric chelates in iron uptake by soybeans. *Plant Physiology*, *50*(2), 208–213. https://doi.org/10.1104/pp.50.2.208

Chen, Q., Yang, C. J., York, A. M., Xue, W., Daskalska, L. L., DeValk, C. A., Krueger, K. W., Lawton, S. B., Spiegelberg, B. G., Schnell, J. M., Neumeyer, M. A., Perry, J. S., Peterson, A. C., Kim, B., Bergstrom, L., Yang, L., Barber, I. C., Tian, F., & Doebley, J. F. (2019). TeoNAM: A Nested Association Mapping Population for Domestication and Agronomic Trait Analysis in Maize. *Genetics*, *213*(3), 1065–1078. https://doi.org/10.1534/genetics.119.302594

Chuang, J. H., & Li, H. (2004). Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome. *PLoS Biology*, *2*(2), E29–E29. PubMed. https://doi.org/10.1371/journal.pbio.0020029

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1), 169–196. https://doi.org/10.1007/s10681-005-1681-5

Collett, D., & Lewis, T. (1976). The Subjective Nature of Outlier Rejection Procedures. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *25*(3), 228–237. JSTOR. https://doi.org/10.2307/2347230

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819. https://doi.org/10.1126/science.1231143

Cuin, T. A., Betts, S. A., Chalmandrier, R., & Shabala, S. (2008). A root's ability to retain K+ correlates with salt tolerance in wheat. *Journal of Experimental Botany*, *59*(10), 2697–2706. https://doi.org/10.1093/jxb/ern128

Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S. L., Briat, J.-F., & Walker, E. L. (2001). Maize yellow stripe1 encodes a membrane protein directly involved in Fe(III) uptake. *Nature*, *409*(6818), 346–349. https://doi.org/10.1038/35053080

Davies, L., & Gather, U. (1993). The Identification of Multiple Outliers. *Journal of the American Statistical Association*, *88*(423), 782–792. https://doi.org/10.1080/01621459.1993.10476339

de Dorlodot, S., Forster, B., Pagès, L., Price, A., Tuberosa, R., & Draye, X. (2007). Root system architecture: Opportunities and constraints for genetic improvement of crops. *Trends in Plant Science*, *12*(10), 474–481.

De Wet, J. M. J., Oestry-Stidd, L. L., & Cubero, J. I. (1979). Origins and evolution of foxtail millets (*Setaria italica*). *Journal d'agriculture Traditionnelle et de Botanique Appliquée*, *26*(1), 53–64.

Dekker, J. (2004). Evolutionary Biology of the Foxtail (*Setaria*) Species-Group. In Inderjit (Ed.), *Weed Biology and Management* (pp. 65–113). Springer Netherlands. https://doi.org/10.1007/978-94-017-0552-3_4

Des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (2013). Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 5–29. https://doi.org/10.1146/annurev-ecolsys-110512-135806

Doebley, J. (1992). Mapping the genes that made maize. *Trends in Genetics*, *8*(9), 302–307. https://doi.org/10.1016/0168-9525(92)90261-2

Dong, Z., Xiao, Y., Govindarajulu, R., Feil, R., Siddoway, M. L., Nielsen, T., Lunn, J. E., Hawkins, J., Whipple, C., & Chuck, G. (2019). The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. *Nature Communications*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-11774-w

Doust, A. N., Devos, K. M., Gadberry, M. D., Gale, M. D., & Kellogg, E. A. (2004). Genetic control of branching in foxtail millet. *Proceedings of the National Academy of Sciences*, *101*(24), 9045–9050. https://doi.org/10.1073/pnas.0402892101

Driks, A., & Eichenberger, P. (2016). *The Bacterial Spore: From Molecules to Systems*. John Wiley & Sons.

Durrett, T. P., Gassmann, W., & Rogers, E. E. (2007). The FRD3-Mediated Efflux of Citrate into the Root Vasculature Is Necessary for Efficient Iron Translocation. *Plant Physiology*, *144*(1), 197. https://doi.org/10.1104/pp.107.097162

Dwivedi, S. L., Upadhyaya, H. D., Senthilvel, S., Hash, C. T., Fukunaga, K., Diao, X., Santra, D., Baltensperge, D., & Prasad, M. (2012). *Millets: Genetic and genomic resources*.

Eda, M., Izumitani, A., Ichitani, K., Kawase, M., & Fukunaga, K. (2013). Geographical variation of foxtail millet, *Setaria italica* (L.) P. Beauv. Based on rDNA PCR–RFLP. *Genetic Resources and Crop Evolution*, *60*(1), 265–274. https://doi.org/10.1007/s10722-012-9832-8

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*(397), 171–185. JSTOR. https://doi.org/10.2307/2289144

Efron, B., & Narasimhan, B. (2020). The automatic construction of bootstrap confidence intervals. *Journal of Computational and Graphical Statistics*, 1-12.

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statist. Sci.*, *1*(1), 54–75. https://doi.org/10.1214/ss/1177013815

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).

Eudy, D. M. (2016). Genetic control of salt tolerance traits in seashore paspalum *(Paspalum vaginatum Sw.)*. [PhD Dissertation] University of Georgia.

Evans, J. R. (1983). Nitrogen and Photosynthesis in the Flag Leaf of Wheat (*Triticum aestivum* L.). *Plant Physiology*, *72*(2), 297–302. https://doi.org/10.1104/pp.72.2.297

Fan, C., Chen, Y., & Long, M. (2008). Recurrent tandem gene duplication gave rise to functionally divergent genes in Drosophila. *Molecular Biology and Evolution*, *25*(7), 1451–1458. PubMed. https://doi.org/10.1093/molbev/msn089

Feldman, M. J., Paul, R. E., Banan, D., Barrett, J. F., Sebastian, J., Yee, M.-C., Jiang, H., Lipka, A. E., Brutnell, T. P., Dinneny, J. R., Leakey, A. D. B., & Baxter, I. (2017). Time dependent genetic analysis links field and controlled environment phenotypes in the model C4 grass *Setaria*. *PLOS Genetics*, *13*(6), e1006841. https://doi.org/10.1371/journal.pgen.1006841

Food and Agriculture Organization of the United Nations. (1973). Calcerous Soils, FAO Soils Bulletin 21. Retrieved from http://www.fao.org/3/x5868e/x5868e00.htm.

Heuzé V., Tran G., Sauvant D., Bastianelli D., Lebas F., 2015. *Foxtail millet (Setaria italica), grain*. Feedipedia, a programme by INRA, CIRAD, AFZ and FAO. https://www.feedipedia.org/node/725 *Last updated on May 11, 2015, 14:34*

Gaudin, A. C., McClymont, S. A., Soliman, S. S., & Raizada, M. N. (2014). The effect of altered dosage of a mutant allele of Teosinte branched 1 (tb1-ref) on the root system of modern maize. *BMC Genetics*, *15*(1), 23. https://doi.org/10.1186/1471-2156-15-23

General Technical Report PSW. (1978). Pacific Southwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.

Gibson, G. (2018). Population genetics and GWAS: A primer. *PLOS Biology*, *16*(3), e2005485. https://doi.org/10.1371/journal.pbio.2005485

Global Soil Regions Map | NRCS Soils. (n.d.). Retrieved August 14, 2019, from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/use/?cid=nrcs142p2_054013

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J., Thompson R., (2010). ASReml User Guide. VSN International Ltd, Hemel Hempstead, UK. URL http://www.VSN-Intl. com.

Goron, T. L., & Raizada, M. N. (2015). Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. *Frontiers in Plant Science*, *6*. https://doi.org/10.3389/fpls.2015.00157

Gregersen, P. L., Holm, P. B., & Krupinska, K. (2008). Leaf senescence and nutrient remobilisation in barley and wheat. *Plant Biology*, *10*, 37–49. https://doi.org/10.1111/j.1438-8677.2008.00114.x

Gries, D., Klatt S., & Runge, M. (1998). Copper-deficiency-induced phytosiderophore release in in the calcicole grass *Hordelymus europaeus*. *New Phytologist*, *140*(1), 95–101. https://doi.org/10.1046/j.1469-8137.1998.00250.x

Grillet, L. and Schmidt, W. (2019), Iron acquisition strategies in land plants: not so different after all. *New Phytol*, 224: 11-18. doi:10.1111/nph.16005

Grillet, L., Mari, S., & Schmidt, W. (2014). Iron in seeds – loading pathways and subcellular localization. *Frontiers in Plant Science*, *4*, 535. https://doi.org/10.3389/fpls.2013.00535

Guerinot, M. L., & Yi, Y. (1994). Iron: Nutritious, Noxious, and Not Readily Available. *Plant Physiology*, *104*(3), 815–820. https://doi.org/10.1104/pp.104.3.815

Gupta, H. S., Hossain, F., Muthusamy, V., & Zunjare, R. U. (2019). Marker-Assisted Breeding for Enrichment of Provitamin A in Maize. In *Quality Breeding in Field Crops* (pp. 139–157). Springer.

He, L., Zhang, B., Wang, X., Li, H., & Han, Y. (2015). Foxtail millet: Nutritional and eating quality, and prospects for genetic improvement. *Frontiers of Agricultural Science and Engineering*, *2*(2), 124–133.

Hell, R., & Stephan, U. W. (2003). Iron uptake, trafficking and homeostasis in plants. *Planta*, *216*(4), 541–551. https://doi.org/10.1007/s00425-002-0920-4

Hether, N.H., Olsen, R.A. & Jackson, L.L. (1984) Chemical identification of iron reductants exuded by plant roots, *Journal of Plant Nutrition*, 7:1-5, 667-676, DOI: 10.1080/01904168409363231

Heuwinkel, H., Kirkby, E. A., Bot, J. L., & Marschner, H. (1992). Phosphorus deficiency enhances molybdenum uptake by tomato plants. *Journal of Plant Nutrition*, *15*(5), 549–568. https://doi.org/10.1080/01904169209364340

Hopkins, B. G., Jolley, V. D., & Brown, J. C. (1992). Variable inhibition of iron uptake by oat phytosiderophore in five soybean cultivars. *Journal of Plant Nutrition*, *15*(1), 125–135. https://doi.org/10.1080/01904169209364305

Huang, Y., Hu, Y., & Liu, Y. (2009). Heavy metal accumulation in iron plaque and growth of rice plants upon exposure to single and combined contamination by copper, cadmium and lead. *Acta Ecologica Sinica*, *29*(6), 320–326. https://doi.org/10.1016/j.chnaes.2009.09.011

Iron Deficiency Anaemia Assessment, Prevention, and Control: A guide for programme managers. (2001) World Health Organization (W.H.O.).

Ishimaru, Y., Suzuki, M., Tsukamoto, T., Suzuki, K., Nakazono, M., Kobayashi, T., Wada, Y., Watanabe, S., Matsuhashi, S., Takahashi, M., Nakanishi, H., Mori, S., & Nishizawa, N. K. (2006). Rice plants take up iron as an Fe3+-phytosiderophore and as Fe2+. *The Plant Journal*, *45*(3), 335–346. https://doi.org/10.1111/j.1365-313X.2005.02624.x

Itanna, F., & Coulman, B. (2003). Phytoextraction of Copper, Iron, Manganese, and Zinc from Environmentally Contaminated Sites in Ethiopia, with Three Grass Species. *Communications in Soil Science and Plant Analysis* (Vol. 34). https://doi.org/10.1081/CSS-120017419

Jacobs, K.A., MacDonald, J. D., Berry, A. M., Costello, L. R. (1997). Rooting Responses of Three Oak Species to Low Oxygen Stress. In: Pillsbury, N.H., Verner, J., Tietje, W. D., technical coordinators. 1997. Proceedings of a symposium on oak woodlands: ecology, management, and urban interface issues;

19–22 March 1996; San Luis Obispo, CA. Gen. Tech. Rep. PSW-GTR-160. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture; p. 91-100

Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S., & Venkataraman, G. (2018). CRISPR for Crop Improvement: An Update Review. *Frontiers in Plant Science*, *9*, 985–985. PubMed. https://doi.org/10.3389/fpls.2018.00985

Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., & Prasad, M. (2019). Genome-Wide Association Study of Major Agronomic Traits in Foxtail Millet (*Setaria italica* L.) Using ddRAD Sequencing. *Scientific Reports*, *9*(1), 5020. https://doi.org/10.1038/s41598-019-41602-6

Jeong, K., Julia, C. C., Waters, D. L. E., Pantoja, O., Wissuwa, M., Heuer, S., Liu, L., & Rose, T. J. (2017). Remobilisation of phosphorus fractions in rice flag leaves during grain filling: Implications for photosynthesis and grain yields. *PLOS ONE*, *12*(11), e0187521. https://doi.org/10.1371/journal.pone.0187521

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Chai, Y., Yang, L., Liu, K., Lu, H., Zhu, C., Lu, Y., Zhou, C., Fan, D., Weng, Q., Guo, Y., Huang, T., Zhang, L., Lu, T., … Han, B. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genetics*, *45*, 957.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *ELife*, *2*, e00471. https://doi.org/10.7554/eLife.00471

Jones, A., Breuning-Madsen, H., Brossard, M., Dampha, A., Deckers, J., Dewitte, O., Gallali, T., Hallett, S., Jones, R., Kilasara, M., Le Roux, P., Michéli, E., Montanarella, L., Spaargaren, O., Thiombiano, L., Van Ranst, E., Yemefack, M., Zougmore, R., (eds.), 2013. Soil Atlas of Africa. European Commission, Publications Office of the European Union, Luxembourg. 176 pp. ISBN 978-92-79-26715-4, doi 10.2788/52319

Josephs, E. B., Stinchcombe, J. R., & Wright, S. I. (2017). What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytologist*, *214*(1), 21–33. https://doi.org/10.1111/nph.14410

Jurkevitch, E., Hadar, Y., Chen, Y., Chino, M., & Mori, S. (1993). Indirect utilization of the phytosiderophore mugineic acid as an iron source to rhizosphere fluorescent *Pseudomonas*. *Biometals*, *6*(2), 119–123.

Kabir, A. H., Khatun, M. A., Hossain, M. M., Haider, S. A., Alam, M. F., & Paul, N. K. (2016). Regulation of Phytosiderophore Release and Antioxidant Defense in Roots Driven by Shoot-Based Auxin Signaling Confers Tolerance to Excess Iron in Wheat. *Frontiers in Plant Science*, *7*, 1684. https://doi.org/10.3389/fpls.2016.01684

Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, *178*(3), 1709–1723. PubMed. https://doi.org/10.1534/genetics.107.080101

Kawakami, K., Largaespada, D. A., & Ivics, Z. (2017). Transposons As Tools for Functional Genomics in Vertebrate Models. *Trends in Genetics : TIG*, *33*(11), 784–801. PubMed. https://doi.org/10.1016/j.tig.2017.07.006

Kersey, P. J., Allen, J. E., Allot, A., Barba, M., Boddu, S., Bolt, B. J., Carvalho-Silva, D., Christensen, M., Davis, P., Grabmueller, C., Kumar, N., Liu, Z., Maurel, T., Moore, B., McDowall, M. D., Maheswari, U., Naamati, G., Newman, V., Ong, C. K., … Yates, A. (2018). Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research*, *46*(D1), D802–D808. https://doi.org/10.1093/nar/gkx1011

Khasanova, A., Lovell, J. T., Bonnette, J., Weng, X., Jenkins, J., Yoshinaga, Y., Schmutz, J., & Juenger, T. E. (2019). The Genetic Architecture of Shoot and Root Trait Divergence Between Mesic and Xeric Ecotypes of a Perennial Grass. *Frontiers in Plant Science*, *10*, 366. https://doi.org/10.3389/fpls.2019.00366

Kichey, T., Hirel, B., Heumez, E., Dubois, F., & Le Gouis, J. (2007). In winter wheat (*Triticum aestivum* L.), post-anthesis nitrogen uptake and remobilisation to the grain correlates with agronomic traits and nitrogen physiological markers. *Field Crops Research*, *102*(1), 22–32. https://doi.org/10.1016/j.fcr.2007.01.002

Kleyman, M., Sefer, E., Nicola, T., Espinoza, C., Chhabra, D., Hagood, J. S., Kaminski, N., Ambalavanan, N., & Bar-Joseph, Z. (2017). Selecting the most appropriate time points to profile in high-throughput studies. *ELife*, *6*, *e18541*. https://doi.org/10.7554/eLife.18541

Kobayashi, T., & Nishizawa, N. K. (2012). Iron Uptake, Translocation, and Regulation in Higher Plants. *Annual Review of Plant Biology*, *63*(1), 131–152. https://doi.org/10.1146/annurev-arplant-042811-105522

Koevoets, I. T., Venema, J. H., Elzenga, J. Theo. M., & Testerink, C. (2016). Roots Withstanding their Environment: Exploiting Root System Architecture Responses to Abiotic Stress to Improve Crop Tolerance. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.01335

Koike, S., Inoue, H., Mizuno, D., Takahashi, M., Nakanishi, H., Mori, S., & Nishizawa, N. K. (2004). OsYSL2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. *The Plant Journal*, *39*(3), 415–424. https://doi.org/10.1111/j.1365-313X.2004.02146.x

Koren, A., Tirosh, I., & Barkai, N. (2007). Autocorrelation analysis reveals widespread spatial biases in microarray experiments. *BMC Genomics*, *8*, 164–164. PubMed. https://doi.org/10.1186/1471-2164-8-164

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*, 29–29. PubMed. https://doi.org/10.1186/1746-4811-9-29

Krohling, C. A., Eutrópio, F. J., Bertolazi, A. A., Dobbss, L. B., Campostrini, E., Dias, T., & Ramos, A. C. (2016). Ecophysiology of iron homeostasis in plants. *Soil Science and Plant Nutrition*, *62*(1), 39–47. https://doi.org/10.1080/00380768.2015.1123116

Kudo, H., Kudo, K., Uemura, M., & Kawai, S. (2015). Magnesium inhibits cadmium translocation from roots to shoots, rather than the uptake from roots, in barley. *Botany*, *93*(6), 345–351. https://doi.org/10.1139/cjb-2015-0002

Kumar, P., Tewari, R. K., & Sharma, P. N. (2008). Modulation of copper toxicity-induced oxidative damage by excess supply of iron in maize plants. *Plant Cell Reports*, *27*(2), 399–409. https://doi.org/10.1007/s00299-007-0453-1

Kwak, I.-Y., Moore, C. R., Spalding, E. P., & Broman, K. W. (2016). Mapping quantitative trait loci underlying function-valued traits using functional principal component analysis and multi-trait mapping. *G3: Genes, Genomes, Genetics*, *6*(1), 79–86.

Lahner, B., Gong, J., Mahmoudian, M., Smith, E. L., Abid, K. B., Rogers, E. E., Guerinot, M. L., Harper, J. F., Ward, J. M., & McIntyre, L. (2003). Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. *Nature Biotechnology*, *21*(10), 1215–1221.

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Applied and Environmental Microbiology*, *75*(15), 5111–5120. https://doi.org/10.1128/AEM.00335-09

Li, L., Ye, L., Kong, Q., & Shou, H. (2019). A Vacuolar Membrane Ferric-Chelate Reductase, OsFRO1, Alleviates Fe Toxicity in Rice (*Oryza sativa* L.). *Frontiers in Plant Science*, *10*, 700. https://doi.org/10.3389/fpls.2019.00700

Li, P., & Brutnell, T. P. (2011). *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *Journal of Experimental Botany*, *62*(9), 3031–3037. https://doi.org/10.1093/jxb/err096

Li, Y., Wu, S., Cao, Y. *et al.* A phenotypic diversity analysis of foxtail millet (*Setaria italica* (L.) P. Beauv.) landraces of Chinese origin. *Genet Resour Crop Evol* 43, 377–384 (1996). *https://doi.org/10.1007/BF00132958*

Lindsay, W. L., & DeMent, J. D. (1961). Effectiveness of some iron phosphates as sources of phosphorus for plants. *Plant and Soil*, *14*(2), 118–126. JSTOR.

Liu, B. H. (2017). Statistical Genomics: Linkage, Mapping, and QTL Analysis. *CRC Press*. https://doi.org/10.1201/9780203738658

López-Arredondo, D., Leyva-González, M., Alatorre-Cobos, F., & Herrera-Estrella, L. (2013). Biotechnology of nutrient uptake and assimilation in plants. *The International Journal of Developmental Biology*, *57*, 595–610. https://doi.org/10.1387/ijdb.130268lh

Lu, C. C., & Tian, H. (2017). Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth System Science Data*, *9*, 181.

Lu, H., Zhang, J., Liu, K. -b., Wu, N., Li, Y., Zhou, K., Ye, M., Zhang, T., Zhang, H., Yang, X., Shen, L., Xu, D., & Li, Q. (2009). Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proceedings of the National Academy of Sciences*, *106*(18), 7367–7372. https://doi.org/10.1073/pnas.0900158106

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), 823. https://doi.org/10.1126/science.1232033

Marschner, P., & Crowley, D. E. (1998). Phytosiderophores decrease iron stress and pyoverdine production of *Pseudomonas fluorescens* PF-5 (pvd-inaZ). *Soil Biology and Biochemistry*, *30*(10), 1275–1280. https://doi.org/10.1016/S0038-0717(98)00039-X

McCouch, S. R., & Doerge, R. W. (1995). QTL mapping in rice. *Trends in Genetics*, *11*(12), 482–487.

Mersmann, O., Beleites C., Hurling R., Friedman A., Ulrich, J. M. (2019). Accurate Timing Functions.

Meyerowitz, E. M. (2001). Prehistory and History of Arabidopsis Research. *Plant Physiology*, *125*(1), 15–19. https://doi.org/10.1104/pp.125.1.15

Miles, C. & Wayne, M. (2008) Quantitative trait locus (QTL) analysis. *Nature Education* 1(1):208

Moon, K. R., van Dijk, D., Wang, Z., Chen, W., Hirn, M. J., Coifman, R. R., Ivanova, N. B., Wolf, G., & Krishnaswamy, S. (2017). PHATE: A Dimensionality Reduction Method for Visualizing Trajectory Structures in High-Dimensional Biological Data. *BioRxiv*, 120378. https://doi.org/10.1101/120378

Morrissey, J., & Guerinot, M. L. (2009). Iron Uptake and Transport in Plants: The Good, the Bad, and the Ionome. *Chemical Reviews*, *109*(10), 4553–4567. https://doi.org/10.1021/cr900112r

Natural Resources Conservation Service, & United States Department of Agriculture. (2010). *Keys to Soil Taxonomy*. Government Printing Office.

Nakanishi, H., Ogawa, I., Ishimaru, Y., Mori, S., & Nishizawa, N. K. (2006). Iron deficiency enhances cadmium uptake and translocation mediated by the Fe2+ transporters OsIRT1 and OsIRT2 in rice. *Soil Science and Plant Nutrition*, *52*(4), 464–469. https://doi.org/10.1111/j.1747-0765.2006.00055.x

Negishi, T., Nakanishi, H., Yazaki, J., Kishimoto, N., Fujii, F., Shimbo, K., Yamamoto, K., Sakata, K., Sasaki, T., Kikuchi, S., Mori, S., & Nishizawa, N. K. (2002). CDNA microarray analysis of gene expression during Fe-deficiency stress in barley suggests that polar transport of vesicles is implicated in phytosiderophore secretion in Fe-deficient barley roots. *The Plant Journal*, *30*(1), 83–94. https://doi.org/10.1046/j.1365-313X.2002.01270.x

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear statistical models* (Vol. 4). Irwin Chicago.

Noor, M. A. F., Cunningham, A. L., & Larkin, J. C. (2001). Consequences of Recombination Rate Variation on Quantitative Trait Locus Mapping Studies: Simulations Based on the *Drosophila melanogaster* Genome. *Genetics*, *159*(2), 581.

Oburger, E., Gruber, B., Schindlegger, Y., Schenkeveld, W. D. C., Hann, S., Kraemer, S. M., Wenzel, W. W., & Puschenreiter, M. (2014). Root exudation of phytosiderophores from soil-grown wheat. *The New Phytologist*, *203*(4), 1161–1174. PubMed. https://doi.org/10.1111/nph.12868

Odonkor, S., Choi, S., Chakraborty, D., Martinez-Bello, L., Wang, X., Bahri, B. A., Tenaillon, M. I., Panaud, O., & Devos, K. M. (2018). QTL Mapping Combined With Comparative Analyses Identified Candidate Genes for Reduced Shattering in *Setaria italica*. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00918

Paape, T., Zhou, P., Branca, A., Briskine, R., Young, N., & Tiffin, P. (2012). Fine-Scale Population Recombination Rates, Hotspots, and Correlates of Recombination in the *Medicago truncatula* Genome. *Genome Biology and Evolution*, *4*, 726–737. https://doi.org/10.1093/gbe/evs046

Paterson, A. H., Lander, E. S., Hewitt, J. D., Peterson, S., Lincoln, S. E., & Tanksley, S. D. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature*, *335*(6192), 721–726. https://doi.org/10.1038/335721a0

Pavlidis, P., Jensen, J. D., Stephan, W., & Stamatakis, A. (2012). A Critical Assessment of Storytelling: Gene Ontology Categories and the Importance of Validating Genomic Scans. *Molecular Biology and Evolution*, *29*(10), 3237–3248. https://doi.org/10.1093/molbev/mss136

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-147, https://CRAN.R-project.org/package=nlme.

Prasada Rao, K. E., de Wet, J. M. J., Brink, D. E., Mengesha, M. H. (1987). Infraspecific Variation and Systematics of Cultivated *Setaria italica*, Foxtail Millet (Poaceae). *Economic Botany*, *41*(1), 108–116. JSTOR.

Printz, B., Lutts, S., Hausman, J.-F., & Sergeant, K. (2016). Copper Trafficking in Plants and Its Implication on Cell Wall Dynamics. *Frontiers in Plant Science*, *7*, 601. https://doi.org/10.3389/fpls.2016.00601

Provart, N. J., Alonso, J., Assmann, S. M., Bergmann, D., Brady, S. M., Brkljacic, J., Browse, J., Chapple, C., Colot, V., Cutler, S., Dangl, J., Ehrhardt, D., Friesner, J. D., Frommer, W. B., Grotewold, E., Meyerowitz, E., Nemhauser, J., Nordborg, M., Pikaard, C., … McCourt, P. (2016). 50 years of Arabidopsis research: Highlights and future directions. *New Phytologist*, *209*(3), 921–944. https://doi.org/10.1111/nph.13687

Radzki, W., Gutierrez Mañero, F. J., Algar, E., Lucas García, J. A., García-Villaraco, A., & Ramos Solano, B. (2013). Bacterial siderophores efficiently provide iron to iron-starved tomato plants in hydroponics culture. *Antonie Van Leeuwenhoek*, *104*(3), 321–330. https://doi.org/10.1007/s10482-013-9954-9

Rajendran, P., Muthukrishnan, J., & Gunasekaran, P. (2003). Microbes in heavy metal remediation. *IJEB Vol.41(09) [September 2003]*. http://nopr.niscair.res.in/handle/123456789/17153

Rascio, N., & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, *180*(2), 169–181. https://doi.org/10.1016/j.plantsci.2010.08.016

Ratzke, C., & Gore, J. (2018). Modifying and reacting to the environmental pH can drive bacterial interactions. *PLOS Biology*, *16*(3), e2004248. https://doi.org/10.1371/journal.pbio.2004248

Ravindran, G. (1991). Studies on millets: Proximate composition, mineral composition, and phytate and oxalate contents. *Food Chemistry*, *39*(1), 99–107. https://doi.org/10.1016/0308-8146(91)90088-6

Reams, A. B., & Neidle, E. L. (2004). Selection for Gene Clustering by Tandem Duplication. *Annual Review of Microbiology*, *58*(1), 119–142. https://doi.org/10.1146/annurev.micro.58.030603.123806

Romheld, V. (1986). Mobilization of iron in the rhizosphere of different plant species. *Advances in Plant Nutrition*, *2*, 155–204.

Römheld, V. (1991). The role of phytosiderophores in acquisition of iron and other micronutrients in graminaceous species: An ecological approach. *Plant and Soil*, *130*(1), 127–134. https://doi.org/10.1007/BF00011867

Römheld, V., & Marschner, H. (1986). Evidence for a Specific Uptake System for Iron Phytosiderophores in Roots of Grasses. *Plant Physiology*, *80*(1), 175. https://doi.org/10.1104/pp.80.1.175

Römheld, V., & Marschner, H. (1990). Genotypical differences among graminaceous species in release of phytosiderophores and uptake of iron phytosiderophores. *Plant and Soil*, *123*(2), 147–153. https://doi.org/10.1007/BF00011260

Römheld, V., Müller, C., & Marschner, H. (1984). Localization and capacity of proton pumps in roots of intact sunflower plants. *Plant Physiology*, *76*(3), 603–606. https://doi.org/10.1104/pp.76.3.603

Römheld, Volker. (1987). Different strategies for iron acquisition in higher plants. *Physiologia Plantarum*, *70*(2), 231–234. https://doi.org/10.1111/j.1399-3054.1987.tb06137.x

Rothschild, M. F., Hu, Z., & Jiang, Z. (2007). Advances in QTL mapping in pigs. *International Journal of Biological Sciences*, *3*(3), 192–197. PubMed. https://doi.org/10.7150/ijbs.3.192

Rout, G. R., & Das, P. (2009). Effect of Metal Toxicity on Plant Growth and Metabolism: I. Zinc. In E. Lichtfouse, M. Navarrete, P. Debaeke, S. Véronique,

& C. Alberola (Eds.), *Sustainable Agriculture* (pp. 873–884). Springer Netherlands. https://doi.org/10.1007/978-90-481-2666-8_53

Russell, T., & Kurtz, R. (2012). A Comparison of Laboratory-Reared Stock and Captured Fruit Flies *(Drosophila melanogaster)* using Upward Movement, Phototaxic, and Starvation Assays Reveals Significant Behavioral Differences. *The Journal of Experimental Secondary Science.* 2(1), 6.

Salih, H., & Adelson, D. L. (2009). QTL global meta-analysis: Are trait determining genes clustered? *BMC Genomics*, *10*(1), 184. https://doi.org/10.1186/1471-2164-10-184

Salt, D. E. (2004). Update on Plant Ionomics. *Plant Physiology*, *136*(1), 2451. https://doi.org/10.1104/pp.104.047753

Salt, D. E., Baxter, I., & Lahner, B. (2008). Ionomics and the Study of the Plant Ionome. *Annual Review of Plant Biology*, *59*(1), 709–733. https://doi.org/10.1146/annurev.arplant.59.032607.092942

Santana, B. V. N., de Araújo, T. O., Andrade, G. C., de Freitas-Silva, L., Kuki, K. N., Pereira, E. G., Azevedo, A. A., & da Silva, L. C. (2014). Leaf morphoanatomy of species tolerant to excess iron and evaluation of their phytoextraction potential. *Environmental Science and Pollution Research*, *21*(4), 2550–2562. https://doi.org/10.1007/s11356-013-2160-5

Schaaf, G., Ludewig, U., Erenoglu, B. E., Mori, S., Kitahara, T., & Von Wirén, N. (2004). ZmYS1 Functions as a Proton-coupled Symporter for Phytosiderophore- and Nicotianamine-chelated Metals. *Journal of Biological Chemistry*, *279*(10), 9091–9096. https://doi.org/10.1074/jbc.M311799200

Shahzad, Z., Rouached, H., & Rakha, A. (2014). Combating Mineral Malnutrition through Iron and Zinc Biofortification of Cereals. *Comprehensive Reviews in Food Science and Food Safety*, *13*(3), 329–346. https://doi.org/10.1111/1541-4337.12063

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331. https://doi.org/10.3758/PBR.16.2.328

Shakoor, N., Ziegler, G., Dilkes, B. P., Brenton, Z., Boyles, R., Connolly, E. L., Kresovich, S., & Baxter, I. (2016). Integration of Experiments across Diverse Environments Identifies the Genetic Determinants of Variation in Sorghum bicolor Seed Element Composition. *Plant Physiology*, *170*(4), 1989–1998. https://doi.org/10.1104/pp.15.01971

Shannon, L. M. , Chen, Q., Doebley, J. F. (2019). A BC2S3 Maize-Teosinte RIL Population for QTL Mapping. *Maize Genetics Cooperation Newsletter*.

Shannon, L. M. (2012). The Genetic Architecture of Maize Domestication and Range Expansion. [PhD Dissertation] The University of Wisconsin-Madison.

Springer, N. M., Anderson, S. N., Andorf, C. M., Ahern, K. R., Bai, F., Barad, O., Barbazuk, W. B., Bass, H. W., Baruch, K., Ben-Zvi, G., Buckler, E. S., Bukowski, R., Campbell, M. S., Cannon, E. K. S., Chomet, P., Dawe, R. K., Davenport, R., Dooner, H. K., Du, L. H., … Brutnell, T. P. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nature Genetics*, *50*(9), 1282–1288. https://doi.org/10.1038/s41588-018-0158-0

Srivastava, D., Shamim, M., Mishra, A., Yadav, P., Kumar, D., Pandey, P., Khan, N. A., & Singh, K. N. (2019). Introgression of semi-dwarf gene in Kalanamak rice using marker-assisted selection breeding. *Current Science (00113891)*, *116*(4).

Striedter, G. F. (2019). Variation across Species and Levels: Implications for Model Species Research. *Brain, Behavior and Evolution*, *93*(2–3), 57–69. https://doi.org/10.1159/000499664

Studer, A. J., & Doebley, J. F. (2011). Do Large Effect QTL Fractionate? A Case Study at the Maize Domestication QTL teosinte branched1. *Genetics*, *188*(3), 673–681. https://doi.org/10.1534/genetics.111.126508

Sun, J., Luu, N. S., Chen, Z., Chen, B., Cui, X., Wu, J., Zhang, Z., & Lu, T. (2019). Generation and Characterization of a Foxtail Millet (*Setaria italica*) Mutant Library. *Frontiers in Plant Science*, *10*, 369. https://doi.org/10.3389/fpls.2019.00369

Suzuki, M., Nozoye, T., Nagasaka, S., Nakanishi, H., Nishizawa, N. K., & Mori, S. (2016). The detection of endogenous 2'-deoxymugineic acid in olives (*Olea europaea* L.) indicates the biosynthesis of mugineic acid family phytosiderophores in non-graminaceous plants. *Soil Science and Plant Nutrition*, *62*(5–6), 481–488.

Takagi, S., Nomoto, K., & Takemoto, T. (1984). Physiological aspect of mugineic acid, a possible phytosiderophore of graminaceous plants. *Journal of Plant Nutrition*, *7*(1–5), 469–477. https://doi.org/10.1080/01904168409363213

Takuno, S., Terauchi, R., & Innan, H. (2012). The Power of QTL Mapping with RILs. *PLOS ONE*, *7*(10), e46545. https://doi.org/10.1371/journal.pone.0046545

Tanksley, S. D. (1993). Mapping Polygenes. *Annual Review of Genetics*, *27*(1), 205–233. https://doi.org/10.1146/annurev.ge.27.120193.001225

Thomine, S., & Schroeder, J. I. (2013). Plant Metal Transporters with Homology to Proteins of the NRAMP Family. *Landes Bioscience.* https://www.ncbi.nlm.nih.gov/books/NBK6452/

Tsukamoto, T., Nakanishi, H., Uchida, H., Watanabe, S., Matsuhashi, S., Mori, S., & Nishizawa, N. K. (2008). 52Fe Translocation in Barley as Monitored by a Positron-Emitting Tracer Imaging System (PETIS): Evidence for the Direct Translocation of Fe from Roots to Young Leaves via Phloem. *Plant and Cell Physiology*, *50*(1), 48–57. https://doi.org/10.1093/pcp/pcn192

van der Schaar, W., Alonso-Blanco, C., Léon-Kloosterziel, K. M., Jansen, R. C., van Ooijen, J. W., & Koornneef, M. (1997). QTL analysis of seed dormancy in Arabidopsis using recombinant inbred lines and MQM mapping. *Heredity*, *79*(2), 190-200.

van der Waals, J. H., & Laker, M. C. (2008). Micronutrient Deficiencies in Crops in Africa with Emphasis on Southern Africa. In B. J. Alloway (Ed.), *Micronutrient Deficiencies in Global Crop Production* (pp. 201–224). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6860-7_8

Virtanen, P., Gommers, R., Oliphant, T.E. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five Years of GWAS Discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Von Wirén, N., Marschner, H., & Romheld, V. (1996). Roots of Iron-Efficient Maize also Absorb Phytosiderophore-Chelated Zinc. *Plant Physiology*, *111*(4), 1119–1125. PubMed. https://doi.org/10.1104/pp.111.4.1119

Von Wirén, N., Römheld, V., Morel, J. L., Guckert, A., & Marschner, H. (1993). Influence of microorganisms on iron acquisition in maize. *Soil Biology and Biochemistry*, *25*(3), 371–376. https://doi.org/10.1016/0038-0717(93)90136-Y

Wairich, A., de Oliveira, B. H. N., Arend, E. B., Duarte, G. L., Ponte, L. R., Sperotto, R. A., Ricachenevsky, F. K., & Fett, J. P. (2019). The Combined Strategy for iron uptake is not exclusive to domesticated rice (*Oryza sativa*). *Scientific Reports*, *9*(1), 16144. https://doi.org/10.1038/s41598-019-52502-0

Walker, E. L., & Connolly, E. L. (2008). Time to pump iron: Iron-deficiency-signaling mechanisms of higher plants. *Current Opinion in Plant Biology*, *11*(5), 530–535. https://doi.org/10.1016/j.pbi.2008.06.013

Wang, W., Vinocur, B., & Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: Towards genetic engineering for stress tolerance. *Planta*, *218*(1), 1–14. https://doi.org/10.1007/s00425-003-1105-5

Wang, Z. M., Devos, K. M., Liu, C. J., Wang, R. Q., & Gale, M. D. (1998). Construction of RFLP-based maps of foxtail millet, *Setaria italica* (L.) P. Beauv. *Theoretical and Applied Genetics*, *96*(1), 31–36. https://doi.org/10.1007/s001220050705

Wu, D., Yamaji, N., Yamane, M., Kashino-Fujii, M., Sato, K., & Ma, J. F. (2016). The HvNramp5 Transporter Mediates Uptake of Cadmium and Manganese, But Not Iron. *Plant Physiology*, *172*(3), 1899–1910. https://doi.org/10.1104/pp.16.01189

Würschum, T., & Kraft, T. (2014). Cross-validation in association mapping and its relevance for the estimation of QTL parameters of complex traits. *Heredity*, *112*(4), 463–468. https://doi.org/10.1038/hdy.2013.126

Xiong, H., Kakei, Y., Kobayashi, T., Guo, X., Nakazono, M., Takahashi, H., Nakanishi, H., Shen, H., Zhang, F., & Nishizawa, N. K. (2013). Molecular evidence for phytosiderophore-induced improvement of iron nutrition of peanut intercropped with maize in calcareous soil. *Plant, Cell & Environment*, *36*(10), 1888–1902.

Yin, X., Stam, P., Kropff, M. J., & Schapendonk, A. H. C. M. (2003). Crop Modeling, QTL Mapping, and Their Complementary Role in Plant Breeding. *Agronomy Journal*, *95*(1), 90–98. https://doi.org/10.2134/agronj2003.9000

Yon Rhee, S., Wood, V., Dolinski, K., & Draghici, S. (2008). Use and misuse of the gene ontology annotations. *Nature Reviews Genetics*, *9*(7), 509–515. https://doi.org/10.1038/nrg2363

Zhang, C. J., Chen, G. X., Gao, X. X., & Chu, C. J. (2006). Photosynthetic decline in flag leaves of two field-grown spring wheat cultivars with different senescence properties. *South African Journal of Botany*, *72*(1), 15–23. https://doi.org/10.1016/j.sajb.2005.03.002

Zheng, L., Fujii, M., Yamaji, N., Sasaki, A., Yamane, M., Sakurai, I., Sato, K., & Ma, J. F. (2011). Isolation and Characterization of a Barley Yellow Stripe-Like Gene, HvYSL5. *Plant and Cell Physiology*, *52*(5), 765–774. https://doi.org/10.1093/pcp/pcr009

Zhu, C., Yang, J., & Shyu, C. (2017). *Setaria* Comes of Age: Meeting Report on the Second International *Setaria* Genetics Conference. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01562

Ziegler, G., Terauchi, A., Becker, A., Armstrong, P., Hudson, K., & Baxter, I. (2013). Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. *The Plant Genome*, *6*(2).