8-2018

Dissection of Soil Waterlogging Tolerance in Soft Red Winter Wheat using Genomic Approaches

Marlovi Andrea Acuna-Galindo
University of Arkansas, Fayetteville

Dissection of Soil Waterlogging Tolerance in Soft Red Winter Wheat using Genomic Approaches

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Crop, Soil, and Environmental Sciences

by

Marlovi Andrea Acuna Galindo
University Nacional de Colombia
Bachelor of Science in Agronomic Engineering, 2008
University of Arkansas
Master of Science in Crop, Soil, and Environmental Sciences, 2012

August 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Richard Esten Mason, Ph.D.
Dissertation Director

_____	_____
Burt H. Bluhm, Ph. D.	Ainong Shi, Ph. D.
Committee Member	Committee Member
_____	_____
David M. Miller. Ph. D.	Andy Pereira, Ph. D.
Committee Member	Committee Member

Abstract

Genomic methods including genome wide association analysis (GWAS), genomic selection (GS) and RNA-seq allow for faster selection of superior breeding lines and for identification and resolution of candidate genes. A panel of 240 soft red winter wheat (*Triticum aestivum* L.) cultivars and breeding lines were subjected to soil waterlogging stress over two seasons at Stuttgart, AR and St. Joseph, LA, US. Total concentrations of P, K, Ca, Mg, Mn, Fe, Al, B, Cu, Na, S and Zn were determined in wheat shoots post-waterlogging using inductively coupled plasma spectroscopy. Yield components kernel number per spike (KNPS), kernel weight per spike (KWS) and thousand kernel weight (TKW) were measured at plant maturity. Negative correlations between TKW and KWS with aluminum and iron concentrations indicated the impact of elemental toxicity on grain production. A ten-fold cross-validation (CV) analysis and ridge regression BLUP (RR-BLUP) model found GS prediction accuracies (r_{gs}) of micro and macronutrient concentrations to range from r_{gs} = 0.06 to 0.52 and improved as more site-years were included in the analysis. The ratio of genomic to phenotypic prediction accuracy (r_{gs} /$H^{1/2}$) was greater than 0.50 for eight of the twelve elements, indicating the potential for using GS to select for shoot micro and macronutrient concentrations in the absence of phenotypic data. GWAS identified forty-seven highly significant ($p < 0.00001$), twenty-three very significant and consistent ($p < 0.0005$) and eight significant and consistent ($p < 0.001$) marker trait associations (MTA) for the twelve micro and macronutrients measured. Lastly, RNA-seq was used for transcriptome and gene expression analysis under waterlogged and non-waterlogged conditions in wheat cultivars ‘Pioneer Brand 26R61’ and ‘AGS 2000’. Around 300 million pair-end reads were developed, covering approximately 16 Gb of the wheat transcriptome. In total, 64,911 (AGS200) and 60,414 (26R61) were obtained and 58,753 expressed genes were observed across

both cultivars and treatments. Overall, the results of this study have and will enable genomics assisted breeding for waterlogging tolerance within the University of Arkansas Wheat Breeding Program by allowing for selection of materials with reduced micro and macronutrient concentrations in new breeding lines in the absence of phenotypic data.

Acknowledgments

First, I would like to thank my major advisor, Dr. Esten Mason for his encouragement, patience, support and guidance during my graduate studies.

Also, I would like to thank my committee members, Dr. Andy Pereira, Dr. Burton Bluhm, Dr. Ainong Shi and Dr. David Miller for their valued advice and to the CSES department staff for their support.

My gratitude to my fellow Wheat Breeding and Genetics Lab members for their support and good moments, Topher, Ben, Amanda, Dennis, Lucas, Nelly, Diana, Nithya, Ricky, Habib, Jamison, Dylan and Zack, I will always treasure our long days in awesome wheat fields, learning the art of plant breeding and having tons of fun.

Many thanks to Guille who has shared this road with me all these years, thanks for your support and encouragement. To my son Vito whose sweetness and passion for learning are my inspiration. Vito, you are my strength to keep going every day. To my mom Lucerito, without you this would not be possible. Thanks for taking care of my treasure all these years, during busy days, long nights and many travels.

Dedication

To my Vito

Table of Contents

Page

List of Tables

CHAPTER II **Page**

CHAPTER III

CHAPTER IV

List of Tables (Cont.)

List of Figures

CHAPTER II **Page**

CHAPTER III

CHAPTER IV

List of Published/Submitted Papers

Acuna, A., Mason, E., Arguello, Holder, A., N., Lozada, D.N., Moon, D., Miller, R., Brown-Guedira, (2018) Genomic selection prediction accuracies for micro and macronutrient accumulation under waterlogging stress in wheat (*Triticum aestivum* L.) Euphytica, Submitted (Chapter II)

Acuna, A., Mason, E., Arguello, Holder, A., N., Lozada, D.N., Moon, D., Miller, R., Brown-Guedira, (2018) Genome wide association analysis of micro and macronutrient concentrations under waterlogging stress in wheat (*Triticum aestivum* L.). Crop Science, submitted (Chapter III)

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

Impact of waterlogging stress on wheat yield and yield components

Approximately 10 to 15 million hectares of the globally cultivated wheat (*Triticum aestivum* L.) area experience soil waterlogging (Boru et al. 2001; MacKintosh et al. 1994; Sayre. et al. 1994), including 16% of the available crop land in United States (Boyer 1982). Soil waterlogging occurs through saturation of the soil pore spaces resulting in an energy crisis due to decreased oxygen levels at the root-soil interface. At the molecular level this hypoxic condition disrupts the oxidative phosphorylation pathway (Juntawong et al. 2014). Wheat losses from 10% to 15% are common with losses of 50% or greater reported in Western Australia (Dennis et al. 2000).

Studies have shown that waterlogging stress can reduce total wheat grain yield (GY) by 21% to 45% dependent on the timing and duration of the stress (Belford 1981; Cannell et al. 1984; Collaku and Harrison 2002; Musgrave and Ding 1998). Cannell et al. (1984) reported reductions in GY and grain number per ear of 16 and 17%, respectively, when waterlogging occurred before emergence. Malik et al. (2001) reported total shoot dry weight reductions between 67% and 72% in three week old winter wheat plants that were subjected to waterlogging for fourteen days, which also decreased seminal root growth by three fold.

Araki et al. (2012a) reported a reduction in thousand kernel weight (TKW) between 22% and 29% for plants subject to waterlogging stress during jointing and anthesis, respectively. This reduction was due in part to both a shortened grain filling duration and decline in carbohydrate translocation.

In soft red winter wheat (SRWW) the impact of waterlogging on yield and yield components was evaluated by Collaku and Harrison (2002). In this study, the performance of nine wheat genotypes under waterlogging conditions showed mean GY reductions of 41% due in part to reductions in both spike density and kernel number of 41 and 20%, respectively. A

follow up study showed kernel weight, chlorophyll content and tiller number to have moderately high heritabilities of H^2 = 47, 37 and 31 %, respectively, which was greater than GY alone at 25% (Collaku and Harrison 2005).

Effect of waterlogging on the soil environment

Soil physical properties and waterlogging

Soil texture is an important factor determining soil water content, water movement and water distribution through the zone root (Evett 2007) and therefore plant ability to cope with waterlogging stress. For example, in duplex soils water accumulation will start from the bottom and move to the top of the soil profile. This results in seminal roots being exposed to hypoxia stress as soon as the water level in the soil rises and increasing susceptibility. In contrast, in sodic and heavy clay soils the waterlogging effect will move from the top to bottom of the soil profile, thus first affecting adventitious roots located in the upper part of the profile (Setter and Waters 2003). Cannell et al. (1984) reported similar results evaluating winter wheat and barley (*Hordeum vulgare* L.) plants under waterlogging conditions. In this study, oxygen decreased consistently at different depths in clay soils and as such, more time was needed for soil recuperation post-waterlogging. As a result, root tissues growing in clay soils under waterlogging stress experience more time under anaerobic conditions than those grown in sandy loam soils under the same conditions.

Mississippi Delta Alluvium Soils

In the Mississippi River Alluvium, texture and soil composition vary greatly due to differences in site deposition (Scott et al. 1998). As a result, the Mississippi Delta soils are very diverse and fertile (Snipes et al. 2005). Dominant taxonomic orders in this area include Inceptisols, Entisols, Vertisols or Alfisols (Scott et al. 1998). Alfisols and Inceptisols span close to 90% of the

Arkansas Delta region with 61 and 28% of coverage, respectively (Escobar 2002). The dominant soil textures include silt loam, clay, silty clay, silty clay loam and fine sandy loam. In general they include deep soils, with low permeability that are poorly drained (Escobar 2002).

Low oxygen (REDOX potential)

Microbial activity is the driving factor during redox reactions (Munch et al. 1978) and plays an important role in processes such as dissolution-precipitation and adsorption-desorption during the solid phase in soils (Kögel-Knabner et al. 2010). Therefore, oxygen content in waterlogged soils is affected by internal factors such as temperature and biological activity (Belford et al. 1985). During a soil waterlogging event, internal micro-spaces are filled with water, decreasing soil oxygen content (Ponnamperuma 1972). As a result, aerobic microorganisms which utilize oxygen as a principal electron acceptor during the respiration process are required to shift first to facultative and then to completely anaerobic respiration (Kögel-Knabner et al. 2010). Alternative electron acceptors are used (Fiedler et al. 2007), with the principal redox reactions including oxygen (O_2) to oxygen hydroxide (H_2O), nitrate (NO_3) to nitrogen (N_2), manganese $Mn^{+4,\,2}$ to manganese Mn^{+2}, iron Fe^{+3} to iron Fe^{+2}, sulfate (SO_4) to hydrogen sulfide (H_2S) and carbon dioxide (CO_2) to methane (CH_4) (Ponnamperuma 1984). These mineral changes decrease soil redox potential (Eh), altering processes such as denitrification, pH and soil chemistry (Pezeshki and DeLaune 2012). Blackwell (1983) reported a rapid decrease in oxygen levels during periods of high temperature in waterlogged soils. Kralova et al. (1992) reported a reduction in microbial activity due to the increase of Eh values (500 -650 mV) which inhibited the denitrification process carried out by microorganisms (Aulakh et al. 1992). Under waterlogging conditions a reduction in the Eh values are accompanied by an increase in pH values decreasing elemental solubility (Kashem and Singh 2001).

Soil environment under waterlogged conditions

Availability and plant uptake of some microelements increases under soil waterlogging conditions (Setter et al. 2009). Under hypoxia, microorganisms increase ion concentrations of NO^{-2} and Fe^{2+} which results in elemental toxicities that affect plant development (Ponnamperuma 1972). Khabaz-Saberi and Rengel (2010) reported an increase in Fe concentration ranging from 3 to 114% under waterlogging, dependent on soil-type.

The effect of waterlogging on wheat grain yield and micro elemental variation of Fe and Mn was evaluated measuring soil redox potential in sodic Indian soil and the acidic Western Australian soil (Yaduvanshi et al. 2012). A clear relationship was also shown between waterlogging effects on soil redox potential and subsequent increase of Fe and Mn concentration. Ding and Musgrave (1995) evaluated the elemental concentration of eleven elements (Boron (B), Calcium (Ca), Copper (Cu), Potassium (K), Magnesium (Mg), Sodium (Na), Sulfur (S), Zinc (Zn), Iron (Fe), Manganese (Mn) and Phosphorus (P) in 14 wheat root genotypes during a three year experiment. Root samples were evaluated using inductively coupled plasma spectroscopy (ICP) were a significate increase in Fe, Mn and P was observed. Moreover, grain yield reductions between 28 to 49% were observed. Khabaz-Saberi et al. (2005) reported an increase in Fe and Na soil concentrations up to 10 fold while evaluating six different wheat genotypes under waterlogging conditions on acidic soil.

Effect of waterlogging on plant environment

Barrett-Lennard (2003) reported three main effects of waterlogging stress on wheat plants, including: 1) A reduction in shoot and root growth caused by low oxygen availability; 2) Disruption of cell membrane permeability limiting nutrient uptake (Trought and Drew 1980), and; 3) A decrease in stomatal conductance leading to a decrease in net photosynthetic rate in

young wheat leaves (Huang et al. 1994). In addition, sugar concentration has been reported to increase in wheat plants after 10 days of waterlogging stress (Barrett-Lennard et al. 1988). It has also been shown that wheat genotypes with tolerance to mineral toxicities outperform sensitive materials under waterlogging conditions (Khabaz-Saberi et al. 2005).

Hypoxia

Waterlogging conditions alter the diffusion rate of oxygen, reducing its availability by 320,000 times (Armstrong 1979; Lee et al. 2006). A decrease in root oxygen leads to hypoxia and anoxia under partial or complete plant submergence, respectively. Under waterlogging roots are completely submerged (hypoxia) while the shoot remains above water level (Ahmed et al. 2013). Hypoxia results in an energy crisis where the production of ATP (adenosine triphosphate) can be reduced by 65 to 97% in affected tissues (Greenway and Gibbs 2003). As oxygen is the main electron acceptor in the oxidative phosphorylation pathway (Dennis et al. 2000), a lack of oxygen will inhibit photosynthesis, respiration and carbohydrate metabolism which are fundamental processes for plant energy production (Drew 1997). Araki et al. (2012b) reported an increase in root respiration in different wheat genotypes under waterlogging stress, which was related to a decrease in root growth development and lower sugar metabolism in root cells.

Lee et al. (2007) showed several ATP production processes including NADH dehydrogenase and NADH ubiquinone oxidoreductase to be down regulated under hypoxic conditions to maintain ATP production and subserve aerenchyma formation. Alteration of ATP and NAD^+ cofactor (nicotinamide adenine dinucleotide) limits coenzyme synthesis for the glycolysis cycle and redox reactions (Dennis et al. 2000). Under normal conditions, plants will produce two molecules of CO_2, two mol of water and 38 mol of ATP through the glucose pathway. In contrast, only two

mol of CO_2, two of ethanol (C_2H_6O) and two of ATP are produced under waterlogging conditions, reducing energy production by as much as 95% (Barrett-Lennard 2003).

Plant adaptations to waterlogging

Classical responses to flooding and soil waterlogging

Plants employ two classical physiological responses to flooding conditions; the low oxygen escape strategy (LOES) and the low oxygen quiescence strategy (LOQS). During LOES oxygen supply is increased by the development or modification of morphological and anatomical traits, including faster elongation of stems and development of aerenchyma in tissues directly affected by the stress (Bailey-Serres and Voesenek 2008). Aerenchyma is formed by longitudinal gas spaces that allows movement of oxygen and other gasses such as ethylene and methane by reducing diffusion resistance in cell tissues (Armstrong 1979). Depending on the mechanism by which it is developed, aerenchyma can be schizogenous or lysigenous. Formation of lysigenous aerenchyma has been shown in wheat, barley, rice (*Oryza sativa* L.) and maize (*Zea Mays* L.) plants (Evans 2003) and can be triggered by abiotic stresses such as mineral deficiencies, waterlogging and low oxygen (Drew and Lynch 1980). In wheat, the development of aerenchyma tissue (Armstrong 1979), and metabolic modification for energy production under low oxygen concentration (Braendle and Crawford, 1987) are two of the principal adaptations involved in waterlogging stress tolerance (Setter and Waters 2003).

In contrast to the LOES, LOQS involves a restriction in developmental growth and cellular metabolism in an effort to diminish energy consumption (Bailey-Serres et al. 2012). In situations where LOES fails due to the impossibility of plant tissues to reach air for oxygen exchange, LOQS can provide longer plant survival (Perata and Voesenek 2007).

Tolerance to micronutrient accumulation

Micro-elemental toxicities have been associated with waterlogging tolerance and variability in tolerance across environments has been attributed to differences in genotypes and soil structure and composition leading to increased concentration or availability of a particular nutrient (Khabaz-Saberi et al. 2005). Metal-protein interactions can cause structural and metabolic damage to cells resulting from an increase in reactive oxygen species (ROS) and oxidative stress (Sharma and Dietz 2009).

Tolerance to aluminum

Aluminum is a micronutrient generally present in all soils in concentrations of 10,000 to 300,000 mg Al kg^{-1}(Lindsay 1979). In acidic soils however, the abundance of Al^{3+} increases and can result in mineral toxicity (Rout et al. 2001). Tolerance of wheat to Al^{3+} toxicity has been associated with the locus *Alt1*, which results in exclusion of Al from the roots and improved root growth (Delhaize et al. 1993). Other mechanisms of Al tolerance include selective permeability in cell walls, immobilization at the cellular level or release of organic compounds for metal inactivation (Delhaize and Ryan 1995; Taylor 1991).

Tolerance to sodium

Salinity stress has been associated with soil waterlogging conditions and acidic soils. Barrett-Lennard. et al. (1999) reported an increase in Na^+ and Cl^- concentrations in wheat shoots of 360 to 650% and 110 to 170%, respectively, under waterlogging. The authors proposed aerenchyma formation, a reduction in stomatal conductance and salt removal strategies such as salt secreting glands, xylem salt absorption and bladder cells as possible mechanisms for salt stress adaptation.

Tolerance to manganese

Manganese toxicity is a major constraint to crop production in waterlogged soils (Foy 1984). Although roots are the main organ involved in Mn absorption, the capacity of the plant to tolerate high shoot Mn concentrations determines tolerance level (Horst 1988). During Mn toxicity, concentrations of both Mn^{3+} and Mn^{4+} increase, leading to the accumulation of phenolic compounds and the disruption of photosystem II (Socha and Guerinot 2014). Genes involved in metal transport have been studied to identify Mn tolerance mechanism in *Arabidopsis thaliana* (Kochian et al. 2004). The expression of the antiporter *CAX2* (calcium exchange 2) was identified as a key factor involved in modifying vacuolar exchange activity to increase plant ability to uptake Ca^{+} and Mn^{+} (Hirschi et al. 2000). Delhaize et al. (2003) reported an increase in Mn tolerance in transgenic *A. thaliana* plants due to expression of the gene *ShMTP1*which increased Mn sequestration to other organelles allowing the plant to store high concentrations of Mn without affecting plant development.

Wheat genome

Common bread wheat is an allohexaploid with its genome evolving from three different species (Chenuil et al. 1999; Mukai et al. 1993). Each of the three wheat genomes is composed of seven chromosomes (1-7), resulting in three homeologous chromosomes across the three genomes (Gill et al. 2004). The A, B and D genomes of *Triticum* aestivum originated from related species *Triticum Urartu/Triticum monococcum* (A), *Aegilops speltoides* (B), and *Aegilops tauschii (D)* (Mukai et al. 1993). Sympatry between the genomes of emmer (AABB) and *A. tauschii* (genomes DD) ultimately resulted in hexaploid wheat (Dubcovsky and Dvorak 2007). The estimated genome size for hexaploid wheat is 16,979 Mbp (Bennett and Smith 1976). The genome is approximately five times the size of the human genome and 45 times larger than the

sequenced genome of rice, making it one of the largest and most complex crop genomes among cultivated species (Arumuganathan and Earle 1991).

Recently, the pre-publication of the reference genome variety *Chinese Spring* was released by the International Wheat genome Sequencing Consortium (IWGSC RefSeq v1.0) https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations. Compilation of previous sequencing efforts in which coding DNA (cDNA), expressed sequence tags (ESTs), bacterial artificial chromosomes sequences (BCAs) and information for physical maps for all 21 chromosomes are included. Mayer et al. (2014) used aneuploidy wheat lines and flow-cytometric sorting to produce a chromosome based draft sequence of hexaploid wheat, an important step towards a full genome sequence. Although only physical map for chromosome 3B is completed (Paux et al. 2008) both physical maps and survey sequences have been completed for nearly all of the 21 wheat chromosomes (http://www.wheatgenome.org/). Additionally, reference genome of the wheat progenitors A and D sub genomes *Aegilops tauschii* (D genome) (Luo et al. 2017) and *Triticum urartu* (A genome) (Ling et al. 2018) have been released.

Molecular markers

Molecular markers are "tags" statistically associated with key agronomic and physiological traits in chromosomal regions of an organism (Collard et al. 2005). Many different types of molecular markers are used in breeding and genetic studies including single nucleotide polymorphism markers (SNPs). SNPs are changes in a unique, single nucleotide in a sequence and are widely present within the wheat genome. Recent development of high-throughput genotyping platforms, including the 9K iSelect Beadchip Assay (Cavanagh et al. (2013) and Illumina 90K platform Wang et al. (2014), in addition to genotype by sequencing tools (He et al. 2014) have expedited the use of molecular markers for dissecting important agronomic traits in wheat. Genotype by

sequencing (GBS) generates high amount of SNP markers at low cost and is considered one of the most cost efficient methods for plant genotyping (Milczarski et al. 2016).

Association Mapping

Complex traits including GY and tolerance to abiotic stresses are influenced by both genetics and the environment and genetic by environmental interactions (GxE). Being quantitatively inherited, they are often controlled by multiple genes with small effects (Ross-Ibarra et al. 2007). Association mapping (AM) evaluates the association between genotypic and phenotype variation within a population of individuals (Yu and Buckler 2006). For AM, development of a bi-parental population is not required as AM is based on historical genetic recombination and linkage disequilibrium (LD) (Lewontin and Kojima 1960; Van Inghelandt et al. 2011). The use of a diverse population of individuals makes it possible to evaluate multiple traits and potentially identify multiple loci at a higher resolution compared to the narrow genetic variation present in LM populations (Edae et al. 2014). In addition, genome coverage is often higher for AM compared to LM as there are more polymorphic markers present in the population (Breseghello and Sorrells 2006).

Two possible approaches are considered in AM studies, including genome wide association (GWAS) and candidate gene analysis (CG) (Álvarez et al. 2014). GWAS is generally used to identify marker-trait associations (MTA) across the whole genome using a large number of molecular markers (Rafalski 2010). In contrast, CG analysis is used to identify MTA for a particular genome region where the biochemical pathway and/or biological function of a candidate gene is known (Ehrenreich et al. 2009).

In association mapping, the concepts of linkage disequilibrium (LD) decay and population structure must be considered and accounted for (Myles et al. 2009). Linkage disequilibrium is

defined as the non-random association between two loci (Ersoz et al. 2007). The LD decay, or breaking of linkage across the genome, is directly related to the number of historical recombination present within the population, which is influenced by both population size and mode of sexual reproduction (self vs. cross-pollinating). Slow LD decay in self-pollinating species such as wheat results in lower resolution of detected MTA regions but fewer markers needed for GWAS. Population structure, or the genetic relatedness among lines within the population must also be considered to reduce the likelihood of detecting false positives due to relatedness of individuals (Flint-Garcia et al. 2003; Myles et al. 2009).

Recent studies using association mapping have been carried out to understand the genetics of quantitative traits in plants. Edae et al. (2014) used a population of 287 elite spring wheat lines to identify genome regions associated with yield components, morphological traits, normalized difference vegetation index (NDVI) as well as traits related to abiotic stress. Four genome regions on chromosomes 5B, 1AL, 5AL, and 3B were associated with these traits. Likewise, Zanke et al. (2015) used GWAS to identify 12 candidate genes associated with grain weight.

Genomic Selection

Genomic selection (GS) is a statistical method that uses high dense marker data to calculate genomic estimated breeding values (GEBV) through the use of a training population (TP) in which genotypic and phenotypic information has been collected (Jannink et al. 2010). GEBVs are calculated analyzing all markers available in the genome to capture the total genetic variance present in a population (Meuwissen et al. 2001). The backbone of the GS approach is the development of a training population in which a genotyped and phenotyped population is used to develop a training model that is used to calculate all model parameters. Then using this training

model, GEBVs can be calculated for new individuals introduced into the model using solely genotypic data (Meuwissen et al. 2001). A reduction in the marker bias is one of the main advantage of genomic selection, because all marker data is used in the calculation of the GEBV rather than only significant markers, eliminating overestimate affects and facilitates the detection of low heritable traits (Heffner et al. 2010).

To test the accuracy of the model, prediction accuracies can be determined using a cross-validation approach in which Pearson correlations are used to determine the fit of the model. Prediction accuracies are defined as the correlation between the observed phenotypic values and GEBV divided by the square root of the heritability of the trait evaluated (Lorenzana and Bernardo 2009). GS can be achieved through the use of different statistical methods, with the common goal being to estimate GEBV using simultaneous estimation of marker effects in a single run (Heslot et al. 2012).

A number of studies have used GS for wheat improvement. Crossa et al. (2010) compared the efficiency of GS vs pedigree base methods in wheat and maize. He observed an increase in prediction ability values by 7.7 to 35.7 with the introduction of genetic markers showing the efficacy of GS. Similarly, Heffner et al. (2011) showed the over performance of GS when compared with phenotypic selection. In their study, a cross-validation approach was used to evaluate 13 agronomic traits on 347 winter wheat lines with 1158 markers. As a result, average prediction accuracies were 28% higher than traditional marker assisted selection and showed an accuracy of 95% higher when compared with phenotypic selection. Grain yield prediction accuracies between 0.28 to 0.45 were observed by Poland et al. (2012) using a GBS data set composed of 41,371 SNP markers and evaluating 254 wheat lines.

Genes associated with waterlogging tolerance and waterlogging conditions

Studies of gene expression are an important tool to identify and understand molecular pathways involved in response to abiotic stresses (Coram 2008). Molecular pathways and genes related to waterlogging and hypoxia stress in plants have been studied previously in different crop species. *Alcohol dehydrogenase (ADH*), a principle enzyme for fermentative metabolism, is expressed at low levels in normal growth conditions in root plants (Chung and Ferl 1999) and overexpression has been shown to be one of the first plant responses to low oxygen supply (Fukao and Bailey-Serres 2004; Sachs et al. 1980). In rice, accumulation of ethylene has been shown to induce the expression of *SNORKEL1 (SK1)* and *SNORKEL2 (SK2)* which are responsible for internode elongation leading to flooding resistance (Hattori et al. 2011). Likewise, Xu et al. (2006) identified allelic variation in the gene *Submergence 1 (Sub1)* with susceptible *sativa* species carrying the allele *Sub1A-2* and the *Sub1A-1* allele providing tolerance. Overexpression of the allele *Sub1A-1* in *O. sativa* spp. *japonica* conferred higher submergence tolerance and triggered the expression of the *ADH* gene. In *A., thaliana* the ethylene response factor (ERF) *AtERF73/HRE1* was shown to be a negative modulator of the ethylene response under both hypoxic and normal conditions. Under hypoxia, both *AtERF73/HRE1* and the ethylene precursor *1-aminocyclopropane-1-carboxylic acid (ACC)* are highly expressed (Yang et al. 2011). The ACC ethylene precursor has also been implicated in signaling response in tomato plants under waterlogging/anoxic conditions (Bradford and Yang 1980). Furthermore, the *AtERF73/HRE1* gene exhibits a similar gene function as the previously mentioned *Sub1*, *SK1* and *SK2* in rice (Yang et al. 2011). In maize, genes involved in aerenchyma formation, reactive oxygen species and Ca^{+} signaling under hypoxia have been shown to be regulated by ethylene (Rajhi et al. 2011).

Transcriptome analysis

Massively parallel sequencing of RNA, known as RNA-seq, is a powerful tool for exploring gene expression (Morozova et al. 2009; Wang et al. 2009; Yang et al. 2015). Previous to RNA-seq, microarray, qPCR and later the serial analysis of gene expression (SAGE) were the main methods used for mRNA studies. Shortcomings of previous technologies have been overcome with RNA-seq (Nagalakshmi 2008) with the key advantage being the ability conduct a transcriptome study without previous knowledge of the genome sequence, making it particularly useful for non-model organisms (Wang et al. 2009). In addition to expression analysis, RNA-seq can provide information on short exon sequences, alternative transcription start sites, alternative splicing variants, untranslated transcription regions (UTRs) and novel genes, making it a powerful tool for wheat gene RNA profiling (Duan et al. 2012; Haas et al. 2002).
Previous transcriptome studies in wheat include the development of 73,521 abiotic stress related ESTs identifying common molecular pathways between stresses (Houde et al. 2006). Laudencia-Chingcuanco (2011) identified 12,901 cold induced genes expressed at different temperature ranges and developmental stages with the *Vrn*-A1 locus found to play an important role for longevity of gene expression but not initial response. Cantu et al. (2011) utilized transgenic wheat lines to show that the upregulation of *grain protein content (GPC)* were of particular importance in early monocarpic senescence and nutrient remobilization.

Few studies have utilized either genomic selection or association mapping approaches to better understand waterlogging tolerance in wheat, nor has RNA-seq been used to characterize gene expression. Therefore, the overall objective of this research is to use three genomics approaches (association mapping, genomic selection and RNA-seq analysis) in order to dissect and understand the waterlogging tolerance mechanisms and stress response pathways of wheat.

Approach of the current Study

Although waterlogging stress is an important constraint for wheat production, few studies have been performed to identify tolerant germplasm and genomic regions. Identification of genes and genome regions associated with tolerance to soil waterlogging will allow a better understanding of this constraint and the improvement of this quantitative trait. Therefore, the overall objective of this research was to evaluate three different approaches to dissect the waterlogging tolerance mechanisms and stress response pathways of wheat. The specific objectives were as follows:

Objective 1. Determine the genomic prediction accuracy for wheat micro and macronutrient concentrations under waterlogging stress using a cross-validation approach. A training population panel (TP) consisting of 240 soft-red winter wheat lines were evaluated to determine prediction accuracies for twelve micro and macronutrient concentrations under waterlogging conditions. The working hypothesis was that genomic prediction accuracies could outperform phenotypic selection.

Objective 2. Identify marker-trait associations for wheat micro and macronutrient concentrations under waterlogging stress using an association mapping approach. A panel of 240 soft-red winter wheat lines were evaluated to identify single nucleotide polymorphism markers associated with micro and macronutrient concentrations, including potential toxicities and deficiencies, in both waterlogging and non-waterlogging treatments in the field. The working hypothesis was that there is significant genetic variation within the soft red winter wheat germplasm to identify SNP markers associated with waterlogging tolerance.

Objective 3. **Identify genes that are differentially expressed under conditions of soil waterlogging using RNA-sequencing.** Wheat cultivars ‘Pioneer 26R61’ and ‘AGS2000’ were used to analyze the hexaploid wheat transcriptome under waterlogging and non-waterlogging

conditions using RNA-seq. The working hypothesis was that there is genetic variation between the cultivars ‘Pioneer 26R61’ and ‘AGS2000’ to identify differentially expressed genes under waterlogging and non- waterlogging conditions.

This all-encompassing approach will lead to the identification of important genome regions and molecular markers associated with waterlogging tolerance and the dissection of biological pathways conditioning the abiotic stress tolerance response and structure of important genes.

References

Ahmed F, Rafii MY, Ismail MR, Juraimi AS, Rahim HA, Asfaliza R, Latif MA (2013) Waterlogging tolerance of crops: breeding, mechanism of tolerance, molecular approaches, and future prospects. BioMed research international 2013:963525

Álvarez MF, Mosquera T, Blair MW (2014) The use of association genetics approaches in plant breeding. plant breeding reviews, Plant Breeding Reviews 38:17

Araki H, Hamada A, Hossain MA, Takahashi T (2012a) Waterlogging at jointing and/or after anthesis in wheat induces early leaf senescence and impairs grain filling. Field Crops Research 137:27-36

Araki H, Hossain MA, Takahashi T (2012b) Waterlogging and hypoxia have permanent effects on wheat root growth and respiration. Journal of Agronomy and Crop Science 198:264-275

Armstrong W (1979) Aeration in higher plants. Academic Press

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208-218

Aulakh MS, Doran JW, Mosier AR (1992) Soil denitrification—significance, measurement, and effects of management. In: Stewart BA (ed) Advances in Soil Science. Springer New York, New York, NY, pp 1-57

Bailey-Serres J, Lee SC, Brinton E (2012) Waterproofing crops: effective flooding survival strategies. Plant Physiology 160:1698-1709

Bailey-Serres J, Voesenek LA (2008) Flooding stress: acclimations and genetic diversity. Annual review of plant biology 59:313-339

Barrett-Lennard E, Leighton P, Buwalda F, Gibbs J, Armstrong W, Thomson C, Greenway H (1988) Effects of growing wheat in hypoxic nutrient solutions and of subsequent transfer to aerated solutions. I. Growth and carbohydrate status of shoots and roots. Functional Plant Biology 15:585-598

Barrett-Lennard EG (2003) The interaction between waterlogging and salinity in higher plants: causes, consequences and implications. Plant and Soil 253:35-54

Barrett-Lennard. EG, Ratingen. Pv, Mathie MH (1999) The developing pattern of damage in wheat (Triticum aestivum L.) due to the combined stresses of salinity and hypoxia: experiments under controlled conditions suggest a methodology for plant selection. Aust J Agric Res, 50:129–136

Belford RK (1981) Response of winter wheat to prolonged waterlogging under outdoor conditions. The Journal of Agricultural Science 97:557-568

Belford RK, Cannell RQ, Thomson RJ (1985) Effects of single and multiple waterloggings on the growth and yield of winter wheat on a clay soil. Journal of the Science of Food and Agriculture 36:142-156

Bennett MD, Smith JB (1976) Nuclear dna amounts in angiosperms. Philosophical transactions of the Royal Society of London Series B, Biological sciences 274:227-274

Blackwell PS (1983) Measurements of aeration in waterlogged soils: some improvements of techniques and their application to experiments using lysimeters. Journal of Soil Science 34:271-285

Boru G, van Ginkel M, Kronstad WE, Boersma L (2001) Expression and inheritance of tolerance to waterlogging stress in wheat. Euphytica 117:91-98

Boyer JS (1982) Plant Productivity and Environment. Science 218:443-448

Bradford KJ, Yang SF (1980) Xylem Transport of 1-Aminocyclopropane-1-carboxylic Acid, an Ethylene Precursor, in Waterlogged Tomato Plants. Plant Physiology 65:322-326

Braendle R, Crafford RMM (1987) Rhizome anoxia tolerance and habitat specialization in wetland species. In Plant life in aquatic and amphibious habitats (ed Crafford, RMM) Special Publication Series of the British Ecological Society (Blackwell Scientific, Oxford, 1987) 5:397- 410

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) Cultivars. Genetics 172:1165-1177

Cannell RQ, Belford RK, Gales K, Thomson RJ, Webster CP (1984) Effects of waterlogging and drought on winter-wheat and winter barley grown on a clay and a sandy loam soil. 1. Crop growth and yield. Plant and Soil 80:53-66

Cantu D, Pearce SP, Distelfeld A, Christiansen MW, Uauy C, Akhunov E, Fahima T, Dubcovsky J (2011) Effect of the down-regulation of the high Grain Protein Content (GPC) genes on the wheat transcriptome during monocarpic senescence. BMC genomics 12:492-492

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proceedings of the National Academy of Sciences 110:8057-8062

Chenuil A, Galtier N, Berrebi P (1999) A test of the hypothesis of an autopolyploid vs. allopolyploid origin for a tetraploid lineage: application to the genus barbus (Cyprinidae). Heredity 82 Pt 4:373-380

Chung H-J, Ferl RJ (1999) Arabidopsis alcohol dehydrogenase expression in both shoots and roots is conditioned by root growth environment. Plant Physiology 121:429-436

Collaku A, Harrison SA (2002) Losses in wheat due to waterlogging approved for publication by the director of the LAES as manuscript no. 01-09-220. Crop Sci 42:444-450

Collaku A, Harrison SA (2005) Heritability of waterlogging tolerance in wheat. Crop Science 45:722-727

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142:169-196

Coram T (2008) Using transcriptomics to understand the wheat genome. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 3

Crossa J, Campos Gdl, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713-724

Delhaize E, Craig S, Beaton CD, Bennet RJ, Jagadish VC, Randall PJ (1993) Aluminum tolerance in wheat (*Triticum aestivum* L.) (I. Uptake and distribution of aluminum in root apices). Plant Physiology 103:685-693

Delhaize E, Kataoka T, Hebb DM, White RG, Ryan PR (2003) Genes encoding proteins of the cation diffusion facilitator family that confer manganese tolerance. Plant Cell 15:1131-1142

Delhaize E, Ryan PR (1995) Aluminum toxicity and tolerance in plants. Plant Physiology 107:315-321

Dennis ES, Dolferus R, Ellis M, Rahman M, Wu Y, Hoeren FU, Grover A, Ismond KP, Good AG, Peacock WJ (2000) Molecular strategies for improving waterlogging tolerance in plants. Journal of experimental botany 51:89-97

Ding N, Musgrave ME (1995) Relationship between mineral coating on roots and yield performance of wheat under waterlogging stress. Journal of experimental botany 46:939-945

Drew M, Lynch J (1980) Soil anaerobiosis, microorganisms, and root function. Annual Review of Phytopathology 18:37 - 66

Drew MC (1997) Oxygen deficiency and root metabolism: Injury and Acclimation Under Hypoxia and Anoxia. Annu Rev Plant Physiol Plant Mol Biol 48:223-250

Duan J, Xia C, Zhao G, Jia J, Kong X (2012) Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC genomics 13:392

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862-1866

Edae E, Byrne P, Haley S, Lopes M, Reynolds M (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theoretical and Applied Genetics 127:791-807

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate Gene Association Mapping of Arabidopsis Flowering Time. Genetics 183:325-335

Ersoz E, Yu J, Buckler E (2007) Applications of linkage disequilibrium and association mapping in crop plants. In: Varshney R, Tuberosa R (eds) Genomics-assisted crop improvement. Springer Netherlands, pp 97-119

Escobar AM (2002) A digital database correlating soil properties to land use in the mississippi delta region of arkansas. University of Arkansas, ProQuest Dissertations Publishing,

Evans D (2003) Aerenchyma formation. New Phytologist 161:35 - 49

Evett SR (2007) Soil water and monitoring technology. Irrigation of Agricultural Crops Agron Monogr 30 (2nd ed), ASA, CSSA, and SSSA,:25–84

Fiedler S, Vepraskas MJ, Richardson JL (2007) Soil redox potential: importance, field measurements, and observations. In: Donald LS (ed) Advances in Agronomy. Academic Press, pp 1-54

Flint-Garcia SA, Thornsberry JM, S E, IV B (2003) Structure of linkage disequilibrium in plants Annual review of plant biology 54:357-374

Foy CD (1984) Physiological effects of hydrogen, aluminum, and manganese toxicities in acid soil. Soil acidity and liming:57-97

Fukao T, Bailey-Serres J (2004) Plant responses to hypoxia--is survival a balancing act? Trends Plant Sci 9:449-456

Gill BS, Appels R, Botha-Oberholster AM, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvorak J, Iwanaga M, Keller B, Li W, McCombie WR, Ogihara Y, Quetier F, Sasaki T (2004) A workshop report on wheat genome sequencing: international genome research on wheat consortium. Genetics 168:1087-1096

Greenway H, Gibbs J (2003) Review: Mechanisms of anoxia tolerance in plants. II. Energy requirements for maintenance and energy distribution to essential processes. Functional Plant Biology 30:999-1036

Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL (2002) Full-length messenger RNA sequences greatly improve genome annotation. Genome biology 3:RESEARCH0029

Hattori Y, Nagai K, Ashikari M (2011) Rice growth adapting to deepwater. Current Opinion in Plant Biology 14:100 - 105

He J, Zhao X, Laroche A, Lu Z-X, Liu H, Li Z (2014) Genotyping by sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science 5

Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic selection accuracy using multifamily prediction models in a wheat breeding program. The Plant Genome 4:65-75

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant breeding with genomic selection: gain per unit time and cost. Crop Science 50:1681-1690

Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic selection in plant breeding: a comparison of models. Crop Science 52:146-160

Hirschi KD, Korenkov VD, Wilganowski NL, Wagner GJ (2000) Expression of arabidopsis CAX2 in Tobacco. Altered metal accumulation and increased manganese tolerance. Plant Physiology 124:125-134

Houde M, Belcaid M, Ouellet F, Danyluk J, Monroy AF, Dryanova A, Gulick P, Bergeron A, Laroche A, Links MG, MacCarthy L, Crosby WL, Sarhan F (2006) Wheat EST resources for functional genomics of abiotic stress. BMC genomics 7:149

Horst W (1988) The physiology of manganese toxicity. Manganese in soils and plants. Springer, pp 175-188

Huang B, Johnson JW, NeSmith DS, Bridges DC (1994) Root and shoot growth of wheat genotypes in response to hypoxia and subsequent resumption of aeration. Crop Science 34:1538-1544

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in functional genomics 9:166-177

Juntawong P, Sirikhachornkit A, Pimjan R, Sonthirod C, Sangsrakru D, Yoocha T, Tangphatsornruang S, Srinives P (2014) Elucidation of the molecular responses to waterlogging in Jatropha roots by transcriptome profiling. Frontiers in Plant Science 5:658

Kashem MA, Singh BR (2001) Metal availability in contaminated soils: I. Effects of flooding and organic matter on changes in Eh, pH and solubility of Cd, Ni and Zn. Nutrient Cycling in Agroecosystems 61:247-255

Khabaz-Saberi H, Rengel Z (2010) Aluminum, manganese, and iron tolerance improves performance of wheat genotypes in waterlogged acidic soils. Journal of Plant Nutrition and Soil Science 173:461-468

Khabaz-Saberi H, Setter TL, Waters I (2005) Waterlogging induces high to toxic concentrations of iron, aluminum, and manganese in wheat varieties on acidic soil. Journal of Plant Nutrition 29:899-911

Kochian LV, Hoekenga OA, Piñeros MA (2004) How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. Annu Rev Plant Biol 55:459-493

Kögel-Knabner I, Amelung W, Cao Z, Fiedler S, Frenzel P, Jahn R, Kalbitz K, Kölbl A, Schloter M (2010) Biogeochemistry of paddy soils. Geoderma 157:1-14

Kralova M, Masscheleyn PH, Patrick jr WH (1992) Redox potential as an indicator of electron availability for microbial activity and nitrogen transformations in aerobic soil. Zentralblatt für Mikrobiologie 147:388-399

Laudencia-Chingcuanco D, Ganeshan S, You F, Fowler B, Chibbar R, Anderson O (2011) Genome-wide gene expression analysis supports a developmental model of low temperature tolerance gene regulation in wheat (*Triticum aestivum* L.). BMC genomics 12:299

Lee TG, Jang CS, Kim JY, Kim DS, Park JH, Kim DY, Seo YW (2006) A Myb transcription factor (*TaMyb1*) from wheat roots is expressed during hypoxia: roles in response to the oxygen concentration in root environment and abiotic stresses. Physiologia Plantarum 129:375-385

Lee TG, Jang CS, Kim JY, Seong RC, Kim IG, Kim DS, Seo YW (2007) Expressed sequence tags from wheat roots under hypoxia. Russian Journal of Plant Physiology 54:659-668

Lewontin R, Kojima K-i (1960) The evolutionary dynamics of complex polymorphisms. Evolution:458-472

Lindsay WL (1979) Chemical equilibria in soils. John Wiley and Sons Ltd.

Ling HQ, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y, Yu Y, Du H, Qi M, Li Y, Lu H, Yu H, Cui Y, Wang N, Chen C, Wu H, Zhao Y, Zhang J, Li Y, Zhou W, Zhang B, Hu W, van Eijk MJT, Tang J, Witsenboer HMA, Zhao S, Li Z, Zhang A, Wang D, Liang C (2018) Genome sequence of the progenitor of wheat A subgenome Triticum urartu. Nature 557:424-428

Lorenzana RE, Bernardo R (2009) Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 120:151-161

Luo MC, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y, McGuire PE, Liu S, Long H, Ramasamy RK, Rodriguez JC, Van SL, Yuan L, Wang Z, Xia Z, Xiao L, Anderson OD, Ouyang S, Liang Y, Zimin AV, Pertea G, Qi P, Bennetzen JL, Dai X, Dawson MW, Muller HG, Kugler K, Rivarola-Duarte L, Spannagl M, Mayer KFX, Lu FH, Bevan MW, Leroy P, Li P, You FM, Sun Q, Liu Z, Lyons E, Wicker T, Salzberg SL, Devos KM, Dvorak J (2017) Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature 551:498-502

MacKintosh C, Lyon GD, MacKintosh RW (1994) Protein phosphatase inhibitors activate anti-fungal defence responses of soybean cotyledons and cell cultures. The Plant Journal 5:137-147

Malik AI, Colmer TD, Lambers H, Schortemeyer M (2001) Changes in physiological and morphological traits of roots and shoots of wheat in response to different depths of waterlogging. Functional Plant Biology 28:1121-1131

Mayer KFX, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Kubaláková M, Číhalíková J, Dubská Z, Vrána J, Šperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, Kobayashi F, Itoh T, Matsumoto T, Sakai H, Tanaka T, Wu J, Ogihara Y, Handa H, Maclachlan PR, Sharpe A, Klassen D, Edwards D, Batley J, Olsen O-A, Sandve SR, Lien S, Steuernagel B, Wulff B, Caccamo M, Ayling S, Ramirez-Gonzalez RH, Clavijo BJ, Wright J, Pfeifer M, Spannagl M, Martis MM, Mascher M, Chapman J, Poland JA, Scholz U, Barry K, Waugh R, Rokhsar DS, Muehlbauer GJ, Stein N, Gundlach H, Zytnicki M, Jamilloux V, Quesneville H, Wicker T, Faccioli P, Colaiacovo M, Stanca AM, Budak H, Cattivelli L, Glover N, Pingault L, Paux E, Sharma S, Appels R, Bellgard M, Chapman B, Nussbaumer T, Bader KC, Rimbert H, Wang S, Knox R, Kilian A, Alaux M, Alfama F, Couderc L, Guilhot N, Viseux C, Loaec M, Keller B, Praud S (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345

Meuwissen T.H.E, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829

Milczarski P, Hanek M, Tyrka M, Stojałowski S (2016) The application of GBS markers for extending the dense genetic map of rye (Secale cereale L.) and the localization of the Rfc1 gene restoring male fertility in plants with the C source of sterility-inducing cytoplasm. Journal of Applied Genetics 57:439-451

Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annual Review of Genomics and Human Genetics 10:135-151

Mukai Y, Nakahara Y, Yamamoto M (1993) Simultaneous discrimination of the three genomes in hexaploid wheat by multicolor fluorescence in situ hybridization using total genomic and highly repeated DNA probes. Genome / National Research Council Canada Genome / Conseil national de recherches Canada 36:489-494

Munch JC, Hillebrand T, Ottow JCG (1978) Transformations in the Feo/Fed ratio of pedogenic iron oxides affected by iron-reducing bacteria. Canadian Journal of Soil Science 58:475-486

Musgrave ME, Ding N (1998) Evaluating wheat cultivars for waterlogging tolerance. Crop Sci 38:90-97

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell Online 21:2194-2202

Nagalakshmi U (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344-1349

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Appels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101-104

Perata P, Voesenek LA (2007) Submergence tolerance in rice requires *Sub1A*, an ethylene-response-factor-like gene. Trends Plant Sci 12:43

Pezeshki SR, DeLaune RD (2012) Soil oxidation-reduction in wetlands and its impact on plant functioning. Biology 1:196-221

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5:103-113

Ponnamperuma FN (1972) The chemistry of submerged soils. Advances in agronomy 24

Ponnamperuma FN (1984) Effects of flooding on soils, Academic Press edn. In : Kozlowski TT, New York

Rafalski JA (2010) Association genetics in crop improvement. Current opinion in plant biology 13:174-180

Rajhi I, Yamauchi T, Takahashi H, Nishiuchi S, Shiono K, Watanabe R, Mliki A, Nagamura Y, Tsutsumi N, Nishizawa NK, Nakazono M (2011) Identification of genes expressed in maize root cortical cells during lysigenous aerenchyma formation using laser microdissection and microarray analyses. The New phytologist 190:351-368

Ross-Ibarra J, Morrell PL, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences of the United States of America 104:8641-8648

Rout GR, Samantaray S, Das P (2001) Aluminium toxicity in plants: a review. Agronomie 21:3-21

Sachs MM, Freeling M, Okimoto R (1980) The anaerobic proteins of maize. Cell 20:761-767

Sayre. KD, Van Ginkel. M, Rajaram. S, Ortiz-Monasterio. I (1994) Tolerance to water-logging losses in spring bread wheat: Effect of time of onset on expression. Annual Wheat Newsl 40:165-171. Colorado State University, Colorado.

Scott D, Ferguson JA, Hanson L, Fugitt T, Smith E (1998) Agricultural water management in the mississippi delta region of arkansas 959

Setter TL, Waters I (2003) Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. Plant and Soil 253:1-34

Setter TL, Waters I, Sharma SK, Singh KN, Kulshreshtha N, Yaduvanshi NP, Ram PC, Singh BN, Rane J, McDonald G, Khabaz-Saberi H, Biddulph TB, Wilson R, Barclay I, McLean R, Cakir M (2009) Review of wheat improvement for waterlogging tolerance in Australia and India: the importance of anaerobiosis and element toxicities associated with different soils. Ann Bot 103:221-235

Sharma SS, Dietz K-J (2009) The relationship between metal toxicity and cellular redox imbalance. Trends in Plant Science 14:43-50

Snipes CE, Nichols SP, Poston DH, Walker TW, Evans LP, Robinson HR (2005) Current agricultural practices of the mississippi delta mississippi. delta research and extension center, Mississippi State University, Stoneville, MS 38776-0197

Socha AL, Guerinot ML (2014) Mn-euvering manganese: the role of transporter gene family members in manganese uptake and mobilization in plants. Frontiers in Plant Science 5:106

Taylor GJ (1991) Current views of the aluminum stress response; the physiological basis of tolerance. Curr Top Plant Biohem Physiol:57-93

Trought M, Drew M (1980) The development of waterlogging damage in young wheat plants in anaerobic solution cultures. Journal of experimental botany 31:1573 - 1585

Van Inghelandt D, Reif JC, Dhillon BS, Flament P, Melchinger AE (2011) Extent and genome-wide distribution of linkage disequilibrium in commercial maize germplasm. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 123:11-20

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol J 12:787-796

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics 10:57-63

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442:705-708

Yaduvanshi NPS, Setter TL, Sharma SK, Singh KN, Kulshreshtha N (2012) Influence of waterlogging on yield of wheat (Triticum aestivum), redox potentials, and concentrations of microelements in different soils in India and Australia. Soil Research 50:489

Yang C-Y, Hsu F-C, Li J-P, Wang N-N, Shih M-C (2011) The AP2/ERF transcription factor AtERF73/HRE1 modulates ethylene responses during hypoxia in arabidopsis. Plant Physiology 156:202-212

Yang W, Yoon J, Choi H, Fan Y, Chen R, An G (2015) Transcriptome analysis of nitrogen-starvation-responsive genes in rice. BMC plant biology 15:1-12

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Current opinion in biotechnology 17:155-160

Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann F, Eichhorn A, Polley A, Jaenecke C, Ganal MW, Röder MS (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (Triticum aestivum L.) by genome-wide association mapping. Frontiers in Plant Science 6:644

CHAPTER II

GENOMIC SELECTION PREDICTION ACCURACIES FOR MICRO AND MACRONUTRIENT ACCUMULATION UNDER WATERLOGGING STRESS IN WHEAT (*TRITICUM AESTIVUM* L.)

Abstract

Tolerance to elemental toxicities is a mechanism to improve crop productivity in waterlogged soils. A training population of 240 winter wheat cultivars was evaluated under field waterlogging conditions to determine shoot concentrations of twelve micro and macronutrients and yield components kernel number per spike (KNPS), kernel weight per spike (KWS) and thousand kernel weight (TKW). Field experiments were conducted at the Rice Research and Extension Center in Stuttgart, Arkansas and the Northeast Louisiana Research Station in St. Joseph, Louisiana over three growing seasons. A ridge regression BLUP model was used to develop genomic selection (GS) prediction models. Shoot micronutrient concentrations varied across locations with heritability ranging from $h^2 = 0.14$ for aluminum to $h^2 = 0.73$ for calcium. Negative correlations between TKW and KWS with aluminum and iron concentrations indicated the impact of elemental toxicity on grain production. Prediction accuracies calculated as the correlation between the phenotypic and the genomic values under a ten-fold cross-validation ranged from $r_{gs} = 0.06$ to 0.52 and improved as more site-years were included in the analysis. The ratio of genomic to phenotypic prediction accuracy, $r_{gs} / H^{1/2}$, was greater than 0.50 for eight of the twelve elements, indicating the potential for using GS to select for shoot micro and macronutrient concentrations. Overall, variation across site-years, genotypes and environmental factors point to the necessity of targeted evaluation of soil waterlogging and elemental toxicity tolerance.

Introduction

Soil waterlogging impacts 12% of agriculture land in the United States, 38% of the Indus Basin in Pakistan as well as large areas in southern Australia and India (Boru et al. 2001; Boyer 1982; Jiang et al. 2008; Kahlown and Azam 2002). In wheat (*Triticum aestivum* L.), grain yield losses up to 50% have been reported in modern cultivars subjected to waterlogging at the vegetative stage, due to a reduction in tiller and kernel number (Arguello et al. 2016). Waterlogging causes hypoxia and anoxia which influences mineral nutrition and micro and macronutrient uptake resulting from biochemical reactions that occur when soil oxygen is depleted (Sairam et al. 2008).

As oxygen depletion occurs there is a progression in the electron acceptors utilized by soil microbes and will generally result in a net increase in iron (II), manganese (II) and aluminum (III), which are all highly available for plant uptake (Stevenson and Cole 1999; Westerman 1987). This results in micronutrient toxicities for manganese, iron (Shabala 2011), sulfur (Colmer and Voesenek 2009) and aluminum (Khabaz-Saberi et al. 2012), in addition to possible deficiencies in nitrogen, phosphorus, potassium, magnesium, copper, zinc and others (Setter et al. 2009; Steffens et al. 2005) which in turn impact grain yield. Manganese toxicity is a global constraint to crop production in waterlogged soils as it interferes with the formation and function of Mg enzymes essential to plant chloroplasts and photosynthesis (Foy 1984; Socha and Guerinot 2014). As a result, it is characterized by a marked decrease in photosynthesis which may be attributed to the disruption of rubisco activity (Houtz et al. 1988). Yaduvanshi et al. (2012) reported a significant reduction of wheat grain yield associated with Mn^{2+} and Fe^{2+} uptake under waterlogged conditions in India.

Tolerance to elemental toxicities is a mechanism to improve productivity in waterlogged soils (Setter et al. 2009). Khabaz-Saberi et al. (2012) showed that wheat genotypes tolerant to high

concentrations of Mn^{2+}, Fe^{2+} and Al^{3+} had improved early root and shoot growth compared to susceptible genotypes. While major genes have been successfully deployed for dry-land scenarios, including for exclusion of Al (*Alt1*) and B (*Bo1*), little progress has been made to date in breeding for improved elemental toxicity tolerance under soil waterlogging (Cai et al. 2008; Schnurbusch et al. 2007). Given the environmental variability and difficulty in phenotyping for micronutrient accumulation and lack of known major loci, Ma et al. (2014) suggested genomic selection (GS) as a potential tool to aid in genetic improvement of micronutrient tolerance.

GS uses genome-wide marker data to predict performance through training and validation of a prediction model (Meuwissen et al. 2001). Genome estimated breeding values (GEBVs) are determined and used to select high-performing individuals when phenotypic data are not available. For accurate determination of GEBVs, the training population must include a representative sample of individuals and alleles of the breeding program in which GS will be applied (Hayes et al. 2009; Heffner et al. 2009; Meuwissen et al. 2001). The benefit of GS is in gain per unit of time if deployed early in the breeding cycle or for traits that are laborious to measure, and lower cost for traits that are costly to measure (Rutkoski et al. 2011). As the aim of GS is to capture all additive effects present in a population, a minimum of one marker in linkage disequilibrium (LD) with each quantitative trait loci (QTL) controlling the trait is necessary for accurate determination of GEBVs (Heffner et al. 2009). This also represents one of the major advantages of GS over marker assisted selection (MAS) methods in which only a limited number of statistically significant markers are used for selection (Heffner et al. 2009; Rutkoski et al. 2011). When phenotypic and genotypic data are available for a training population, a k-fold cross-validation (CV) approach can be used to evaluate model accuracy.

Studies have shown the applicability of GS for wheat improvement, including for quality (Heffner et al. 2011; Kristensen et al. 2018), grain micronutrient content (Manickavelu et al. 2017), yield and agronomic traits (Saint et al. 2016; Thavamanikumar et al. 2015) and disease resistance (Arruda et al. 2015; Hoffstetter et al. 2016; Rutkoski et al. 2011). GS is particularly promising for improving traits with low heritability (Ziyomo and Bernardo 2013).

Here we present the results evaluating a training population of 240 soft winter wheat cultivars and an initial analysis using GS for the prediction of shoot micro and macronutrient concentrations resulting from field soil waterlogging stress. This study will enable deployment of GS to improve wheat productivity in waterlogged soils.

Materials and methods

Germplasm and experimental design

The training population (TP) used for this study consisted of 240 inbred lines of soft red winter wheat, including cultivars and breeding lines from the SUNGRAINS (Southeastern University Grains) small grain breeding and genetics group (http://www.sungrains.lsu.edu/), other US universities and private wheat breeding programs. Data on the panel was previously described by Lozada et al. (2017). Based on molecular markers for height loci *Rht-B1* and *Rht-D1*, the panel is comprised of 207 semi-dwarf lines (homozygous for either *Rht-B1b* or *Rht-D1b*), 26 wild-type (homozygous for both *Rht-B1a* and *Rht-D1a*) and two 'double-dwarf' (homozygous for both having the *Rht-B1b* and *Rht-D1b*) lines.

Field experiments were conducted at the Rice Research and Extension Center (RREC) in Stuttgart (ST) Arkansas, AR and the Northeast Louisiana Research Station St. Joseph (SJ), LA during the 2012-2013 (13) and 2013-2014 (14) growing seasons. Yield components were also evaluated during 2014-2015 (15) growing season at RREC, Stuttgart, AR. Stuttgart soils are

characterized by a silt loam surface layer and a clay subsoil with low permeability (NRCS 2013) and are prone to periodic waterlogging. St. Joseph soils are characterized by a clayey surface layer and are prone to waterlogging (NRCS 2001). The TP was sown in 2012-2013 in a randomized complete block design (RCBD) with two replications (with the exception of SJ14, which had only a single replication) and in an augmented incomplete block design with two repeated check lines ('Jamestown' and 'Pioneer Brand 26R20') and two replications in 2013-2014 and 2014-2015. Experimental plots at both locations were drill seeded at a rate of 115 seed m^{-2} in two 1.5 m rows. In Stuttgart (ST13, ST14 and ST15), plots were fertilized with 170 kg N ha^{-1} as urea in a split application, with 60% applied prior to the waterlogging treatment and 40% applied post waterlogging treatment. In St. Joseph (SJ13 and SJ14), plots were fertilized with 225 kg N ha^{-1} as urea in a split application, with 50% applied prior to the waterlogging treatment and 50% applied post waterlogging treatment.

Waterlogging treatment

At all locations, the waterlogging treatment was imposed by establishing 0.30 m high levees surrounding the experimental field sites. All treatments at all locations were imposed when plants reached Feekes growth stage 4 or 5 and were terminated at Feekes growth stage 5 or 6. In ST13 and ST14 ground water was applied from a nearby irrigation well to saturate the soil twice weekly for the duration of the treatment as previously described (Arguello et al. 2016). In ST13, the waterlogging treatment was initiated on March 20, 2013 and was terminated on April 17, 2013. In ST14 and ST15, the waterlogging treatment began on April 1 and was terminated on April 14. For SJ13, no artificial waterlogging treatment was imposed due to excessive rainfall during the period but levees were still established to contain naturally occurring rainfall. In SJ14, the waterlogging treatment was initiated on Feb 24 and terminated on March 19.

Phenotypic trait measurements

Total concentrations of macronutrients calcium (Ca), potassium (K), magnesium (Mg), phosphorus (P) and sulfur (S) and micronutrients aluminum (Al), boron (B), copper (Cu), iron (Fe), manganese (Mn), sodium (Na) and zinc (Zn) were determined using inductively coupled plasma atomic emission spectroscopy (ICP-AES) according to Donohue and Aho (1992). For ICP, a 0.10 m^2 section of shoot tissue (stems and leaves) per plot was harvested 24 hours after waterlogging treatment was terminated when the plants ranged from Feekes growth stage 5 to 6. After harvest, soil and dust particles were removed using deionized water. Dry weight was obtained after drying for 72 hours at 65˚C and a 0.25g sample of ground and homogenized tissue (0.5-1.0 mm particle size) was digested with concentrated HNO_3 for one hour at 125˚C. After digestion, samples were dried at 80˚C and treated with 30% H_2O_2 according to the Organic Matter Destruction-Wet Ashing protocol (Cambell and Plank 1992). Yield components including kernel number per spike (KNPS) and kernel weight per spike (KWP) were estimated by harvesting 50 spike-bearing culms per plot at maturity. Thousand kernel weight (TKW) was taken as the weight of a thousand kernels counted using a Seedburo® 801 seed counter (Chicago, IL, USA).

Phenotypic data analysis

Phenotypic data were analyzed using PROC MIXED in SAS 9.4 (SAS Institute Inc. 2011, Cary, NC). Data were analyzed in three ways: 1) Each site-year separately due to significant genotype by site-year interaction; 2) By location across years, and; 3) In a combined analysis across all site-years. Site-years ST13, SJ13 and SJ14 were analyzed as a RCBD with genotype considered a fixed effect and replication was regarded random. ST14 was analyzed as an augmented incomplete block design. For the augmented design, genotypes, incomplete blocks, replications

and the interaction of genotype and replication were regarded as random effects. Adjusted means for each genotype per year were estimated using a restricted maximum likelihood (REML) approach for each replication. Best linear unbiased estimators (BLUEs) for each trait and line were estimated for each individual site-year with genotype considered a fixed effect. Best linear unbiased predictors (BLUPs) for each trait across site-years were estimated using the REML method with all effects (genotypes, site-year, replication and genotype x site-year interaction) considered random. Trait narrow sense (assuming no dominant effects) heritability (h^2) was estimated from the variance components using TYPE3 sum of squares and the formula:

$$\boldsymbol{h^2} = \frac{\boldsymbol{\sigma_G^2}}{\boldsymbol{\sigma_G^2} + \boldsymbol{\sigma^2_{\frac{GEI}{e}}} + \boldsymbol{\sigma^2_{\frac{E}{re}}}}$$

where $\boldsymbol{\sigma_G^2}$, $\boldsymbol{\sigma^2_{\frac{GEI}{e}}}$, and $\boldsymbol{\sigma^2_{\frac{E}{re}}}$ are the genotypic, genotype-by-site-year and error variances, respectively, and ***e*** and ***re*** are the numbers of site-years and number of total replications, respectively. For h^2 calculations within a single site-year, the $\sigma^2_{\frac{GEI}{e}}$ component was removed from the equation, which determines h^2 across replications. Pearson correlations (P_r) among elements were calculated using the BLUEs or BLUPs in PROC CORR.

Genotypic data

Genome wide marker data for the TP panel was generated using genotype by sequencing (GBS). DNA was extracted using Mag-Bind® Plant DNA Plus kit from Omega Bio-tek (Norcross, GA, USA), following the manufacturer's instructions. Genomic DNA was quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit and normalized to 20 ngµl^{-1}. GBS libraries were created using *Pst1-Msp1* restriction enzyme combination consistent with Poland et al. (2012). The samples were pooled together at 96-plex to create libraries and each library was sequenced on a single lane of Illumina Hi-Seq 2500. SNP calling was performed using the TASSEL 5 GBSv2

pipeline (https://bitbucket.org/tasseladmin/tassel-5 source/wiki/Tassel5GBSv2Pipeline) (Glaubitz et al. 2014) using 64 base kmer length and minimum kmer count of five. Reads were aligned to wheat reference "IWGSC_WGAv1.1" (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies) using aln method of Burrows-Wheeler aligner (BWA) version 0.7.10 (Li and Durbin 2009). Raw SNP data generated from the TASSEL pipeline was filtered to remove taxa with more than 90% missing data. Genotypic data was then filtered to select for bi-allelic SNPs with minor allele frequency (MAF) ≥ 5%, missing data ≤ 50% and heterozygosity ≤ 10%. After filtering, imputation for missing data was performed using the LD-kNNi method in TASSEL (Money et al. 2015) resulting in a total of 62,372 markers.

Statistical model and cross-validation

Ridge regression BLUP (RR-BLUP) was used for GS predictions using the 'rrBLUP' package v4.4 in R (Endelman 2011). The RR-BLUP model assumes random effects have a normal distribution and a common variance which allows for simultaneous estimation of all marker effects even when marker number is greater than the number of genotypes (Heffner et al. 2009). The 'kin.blup' function was used to solve the mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \varepsilon \quad ,$$

where $\boldsymbol{Y}$ is the vector for the observed phenotypes, $\boldsymbol{X}$ is the design matrix for fixed effects $\boldsymbol{\beta}$; $\boldsymbol{Z}$ is the design matrix for the vector of random marker effects $\boldsymbol{u}$ with $\boldsymbol{u} \sim \boldsymbol{N}(\mathbf{0}, \mathbf{K}\sigma^2_u)$, $\mathbf{K}$ is a positive semidefinite matrix with normal residuals and a constant variance. In this model variance components were estimated by the default REML approach (Endelman 2011).

Prediction accuracies for GS were determined using tenfold cross-validation (CV) following the equation described by Daetwyler et al. (2008):

$$r_{GS} = \sqrt{\frac{h^2}{h^2+1}} \quad ,$$

where r_{GS} is the accuracy calculated as a function of the heritability (h^2) and the ratio of the number of phenotypes used. For CV, 217 lines were assigned to the TP while 23 lines were assigned to validation set. Cross-validation was repeated for 1000 cycles for each trait and model.

Results

Phenotypic within individual site-years

Shoot micro and macronutrient concentrations varied across locations and years, with the 2012-2013 site-years generally having higher concentrations compared to 2013-2014 (Table 1). For example, mean Mn concentrations were 215 ppm in ST13, compared to 140.25 ppm in ST14, 28.02 ppm in SJ13 and 13.37 ppm in SJ14. Significant genotype effects ($P \leq 0.05$) were observed for all traits in ST13 with the exception of B and Fe. In ST14, significant genotype effects ($P \leq 0.05$) were observed for Ca, Mn, P, Na, S and Zn. In SJ13 significant genotype effects ($P \leq 0.05$) were observed for all elements with the exception of Al, B and Fe (Table 1). Heritability within a site-year ranged from 0.01 to 0.76. In Stuttgart, Na was generally the most heritable trait (h^2 = 0.76 and 0.45) followed by Mn (h^2 = 0.44 and 0.44), Zn (h^2= 0.63 and 0.26), S (h^2= 0.67 and 0.38), and Ca (h^2= 0.50 and 0.39) for ST13 and ST14, respectively. In SJ13 Ca (h^2= 0.71) was the most heritable, followed by Mg (h^2= 0.53), P (h^2= 0.58), K (h^2= 0.67) and S (h^2= 0.66) (Table 1).

Nutrient mean concentrations above the high optimum level determined by Plank and Donohue (2000) were observed for Fe (> 301 ppm) in ST13 and SJ13, Mn (> 201 ppm) in ST13 and K (> 3.01 %) in SJ13. Maximum concentrations above the high optimum level were

observed for K (4.88 %) and B (8.42 ppm) in ST13; Fe (1038 ppm) and Mn (295 ppm) in ST14; P (0.69%) and B (9.11 ppm) in SJ13, and K (3.79%), B (7.9 ppm) and Fe (306.1 ppm) in SJ14. In addition, minimum (4.42 ppm) and maximum (2223 ppm) Al concentrations were observed in ST14 and ST13 respectively.

Phenotypic analysis across site years

Significant genotype effects ($P \leq 0.05$) were observed for all micro and macronutrients when analyzed across all four-site years, with the exception of Al (Table 2). Heritabilities were low to moderate but tended to increase with the addition of more site-years to the analysis. Across all environments, h^2 ranged from 0.14 for Al to 0.73 for Ca. When Stuttgart was analyzed separately, h^2 ranged from 0.00 for Cu to 0.59 for Na. For St. Joseph, h^2 ranged from 0.02 for Cu to $h^2 = 0.66$ for Ca (Table 2). In Stuttgart, yield components showed significant ($P < 0.005$) genotype and genotype by year effects with heritability ranging from $h^2 = 0.20$ for KNS to $h^2 = 0.80$ for TWK. (Table 3).

Correlations among traits

Pearson correlations (P_r) between micro and macronutrient concentrations were generally consistent across the ST and SJ locations (Fig. 1). High and consistent correlations ($r > 0.50$) were observed between Al and Fe ($r = 0.88$ and 0.96), S and K ($r = 0.62$ and 0.56) and S and Mg ($r = 0.69$ and 0.52) in ST and SJ, respectively. In addition, high correlations were observed for P and Mn ($r = 0.61$) and Mg and K ($r = 0.56$) in ST. Similarly, high correlations among P and Zn ($r = 0.61$) were observed in SJ. Inconsistent correlations between the two sites included significant positive correlations in ST for Al and K ($r = 0.19$), Fe and K ($r = 0.24$), Cu and K ($r = 0.18$) and P and Na ($r = 0.41$) which were not observed in SJ (Fig. 1). At both sites, Ca was significantly correlated with all micro and macronutrients with the exception of Al and Cu in ST and Cu in SJ.

In ST, Zn was significantly correlated with all micro and macronutrients (r = 0.15 to 0.44), however no correlation of Zn with Cu and Na was observed at the SJ site.

Yield components were weakly but negatively correlated with some micro and macronutrients (Fig 2). TKW was negatively correlated with Al (r = -0.12), B (r = -0.11), Fe (r = -0.13), Mg (r = -0.13) and Mn (r = -0.11). KWS was negatively correlated with Al (r = -0.15), Fe (r = -0.13) and Na (r = - 0.18), while a positive correlation was observed for Ca (r = 0.14). KNPS was positively correlated with Ca (r = 0.11), K (r = 0.12), and P (r = 0.16) (Fig. 2).

GS prediction accuracies for individual site years

Prediction accuracies (r_{gs}) for shoot micro and macronutrient concentrations were low to moderate within individual site years (Table 4). For Stuttgart, mean r_{gs} across all nutrients was 0.23 and 0.16 for ST13 and ST14, respectively. Accuracy for individual elements ranged from r_{gs} = 0.11 to 0.31 in ST13 and from r_{gs} = -0.10 to 0.36 in ST14, with Na having the highest r_{gs} in both years. For St. Joseph, mean r_{gs} across all nutrients was 0.29 and 0.23 in SJ13 and SJ14, respectively. Accuracy for individual elements ranged from r_{gs} = 0.13 to 0.45 in SJ13 and from r_{gs} = -0.06 to 0.42 in SJ14. Calcium and sulfur had the highest prediction accuracy for SJ13 and SJ14, respectively.

Prediction accuracy of combined site-years

Mean r_{gs} of nutrient concentrations increased as more site-years were included for determining the BLUPs used for GS (Table 4). Mean r_{gs} for Stuttgart (STCOMB) was 0.25, compared to 0.23 and 0.16 for ST13 and ST14, respectively. Similar results were observed for St. Joseph (SJCOMB), with a mean of r_{gs} = 0.30 compared to 0.29 and 0.23 for SJ13 and SJ14, respectively. When all site-years were included (STSJALL), mean r_{gs} increased to 0.32. The STSJALL analysis resulted in the highest r_{gs} for all macronutrients, ranging from r_{gs} = 0.30 to

0.52. In contrast, micronutrient predictions reached a maximum in individual site-years or for the case of B and Mn, in the SJCOMB (Table 4).

Comparison of genomic versus phenotypic selection

For individual site-years, phenotypic selection accuracy, determined as the square root of the narrow-sense heritability ($H^{1/2}$), was superior to GS across all traits with the exception of B, where the r_{gs} /$H^{1/2}$ ratio was 1.28 for SJ13. The $r_{gs}/H_{1/2}$ ratio increased when data were combined across years within a site and across all site-years. For both STCOMB and SJCOMB, r_{gs} /$H^{1/2}$ was > 0.50 for six of the 12 nutrient concentrations. For STSJALL, r_{gs} /$H^{1/2}$ was > 0.50 for 8 of the 12 traits (Table 4).

Discussion

Trait variability and Pearson correlations

We evaluated shoot micro and macronutrient concentrations in a training population of 240 SRWW genotypes grown under field waterlogging conditions in Stuttgart, AR and St. Joseph, LA. As expected, there were significant differences for accumulation of shoot micro and macronutrients across years and locations (Table 1). Previous studies have reported the complexity of nutrient interactions under waterlogging conditions (Khabaz-Saberi et al. 2005), as plant nutrient concentrations are driven by soil parent material and soil site history (Ågren and Weih 2012). Although soils at both sites used in this study were prone to waterlogging, screening of breeding material under these conditions is challenging due to a lack of uniformity in soil conditions and stress severity (Chaubey and Senadhira 1994). In our study, there was variability in the treatment duration and intensity due to environmental factors. For example, artificial flooding was not necessary in SJ13 due to heavy precipitation during the treatment period. In ST14 waterlogging stress was imposed for fourteen days (as oppose to 28 days in ST13) due to

unusually warm temperatures during the treatment period and to avoid the premature death of the wheat plants.

Variation across-site years, crop genotypes and environmental factors indicate the necessity of target location evaluations for soil waterlogging and mineral toxicities tolerance. In SJ13, macronutrients had the highest minimum shoot concentrations and showed strong significant genotype effects, which could indicate the requirement of a minimum threshold to identify differential responses among genotypes. Although toxic levels were not achieved for all twelve elements evaluated, they were observed for Fe and Al in at least two of the four site-years. Inhibition of wheat root development at nutrient solution concentrations of 3.2 ppm was reported by Kerridge (1968). In this study, [Al] was above toxic levels in all genotypes evaluated and in all site-years, reaching a high maximum of 2223 ppm in the genotype 'Holley'. The Fe concentrations observed were also above the sufficient level ranges recommended during seedling to tillering in wheat (30 to 200 ppm) (Plank and Donohue 2000) with a maximum high of 2328 ppm for genotype 'NC08-23925'. These results were consistent with Pearson correlation analysis in which a strong ($p < 0.0001$) positive correlation between Al and Fe was observed in all locations. Although previous studies reported increases in Al, Fe and Mn concentrations in acidic soils as main factors for plant decay under waterlogging conditions (Khabaz-Saberi and Rengel 2010; Khabaz-Saberi et al. 2005; Setter et al. 2009), no positive correlations between Al and Fe under waterlogging conditions have been reported.

Manganese concentrations above the recommended sufficient levels during seedling and tillering (20 to 150ppm) were observed in ST13 (Plank C.O. and Donohue. 2000). An increase in [Mn] is associated with an increase in [Al] in acidic soils (Khabaz-Saberi and Rengel 2010) and [Fe] in sodic soils (Yaduvanshi et al. 2012). In this study, the highest shoot concentrations of Al

and Fe were observed in ST13 which could in turn impact [Mn]. Additionally, correlation analysis showed a positive correlation between [Mn] and [Fe] for all years except for ST13, while a negative significant correlation between Mn and [Al] was observed. Under waterlogging conditions, ferric and manganic forms are reduced to ferrous and manganous states increasing Fe and Mn solubility (Gotoh and Patrick 1974; Ponnamperuma 1972) which can increase [Mn] in shoot tissue as reported by Trought and Drew (1980). However, previous studies in rice (*Oryza sativa* L.) have shown that [Al] can also modify the root symplastic pathway, decreasing Mn shoot concentrations but increasing accumulation of Mn in the roots through modification of binding properties of root cell walls (Wang et al. 2015). Our results show the complex interaction among elements under waterlogging conditions where soil chemistry is impacted by several factors such as microbial activity, soil temperature, anoxia conditions, mineral content and soil characteristics (Armstrong 1979; Munch et al. 1978; Ponnamperuma 1972; Setter and Waters 2003).

Although no total grain yield was measured in this study, negative significant correlations among yield components (TKW and KWS) and [Al] and [Fe] suggest a direct impact on grain production under WL conditions, in agreement with (Arguello et al. 2016) in which a reduction of TKW (14%) and KWS (28%) was observed under waterlogging conditions in winter wheat.

Heritability

Evaluation of genotypes across site-years (STSJALL) resulted in higher heritability for most traits by reducing GxE and error variances. Exceptions included K and S, where a decrease in heritability from individual analysis of K ($h^2 > 0.18$) and S ($h^2 > 0.38$) was observed compared to the combined analysis K ($h^2 = 0.2$) and S ($h^2 = 0.18$) due to a strong genotype by environment interaction. Moderate to high heritabilities values obtained in this study, indicate that a portion

of the observed phenotype for the traits evaluated is due to genetic variation and thus molecular markers could be used to select for genetic variation through GS (Kristensen et al. 2018).

Genomic selection for micro and macronutrient concentrations

We used 10 fold cross-validation to determine the prediction accuracy of genomic selection for twelve micro and macronutrient shoot concentrations evaluated in 240 SRWW lines. Low to moderate prediction accuracies were obtained for most of the traits evaluated in this study, ranging from (r_{gs} = -0.10 to 0.50). Cross validation using combined data from all four site-years out performed single site analysis. Increases in prediction accuracies have been observed in multivariate analysis studies where additional information from individual traits was added to the GS model (Lopez-Cruz et al. 2015; Okeke et al. 2017). However, Resende et al. (2012) reported a decrease in prediction accuracies across populations and emphasize the importance of using target environments to increase prediction accuracies. In this study, weakly heritable nutrient concentrations (h^2 < 0.20) had low prediction accuracies (r_{gs} = < 0.15). These results were expected, based on previous studies (Combs and Bernardo 2013; Wong and Bernardo 2008). Therefore, to optimize prediction accuracies for lowly heritable traits, particularly for Al, Fe and Mg in this study, a target environment approach with an increased number of replications must be used.

To the best of our knowledge, only one similar study under non-waterlogging conditions evaluated grain Fe and Zn concentrations in spring wheat using GS. Prediction accuracies for Fe and Zn ranging from (r_{gs} = 0.32 and 0.73) and (r_{gs} = 0.33 and 0.69) were reported (Velu et al. 2016). In our study, lower prediction accuracies were obtained ranging from (r_{gs} = -0.02 to 0.23) and (r_{gs} = 0.18 to 0.39) for Fe and Zn respectively. However, a comparison among studies evaluating accuracy predictions of complex traits such as mineral toxicities are difficult to make,

as reported by Crossa et al. (2010), in which evaluation of the same population in different environments resulted in different prediction accuracies.

Conclusions

This is the first study evaluating micro and macronutrient concentrations in wheat under waterlogging conditions using a GS approach. Although validation of these preliminary results is needed, this study marks an important resource for improving the waterlogging tolerance of wheat. Moderate to high prediction accuracies for Ca, K, Mg, S, Mn, Na and Zn indicate that the prediction models developed in this study could be used for the selection of tolerant lines in the absence of phenotypic data.

Table 1. Descriptive statistics and analysis of variance of macro and micronutrient concentrations in individual site-years for 240 wheat genotypes grown under field waterlogging conditions in Stuttgart AR and St. Joseph LA, 2012-2014.

	Macronutrients					Micronutrients						
Source	**Ca %**	**K %**	**Mg %**	**P %**	**S %**	**Al ppm**	**B ppm**	**Cu ppm**	**Fe ppm**	**Mn ppm**	**Na ppm**	**Zn ppm**
Stuttgart (2012-2013)												
Mean	0.33	2.92	0.11	0.29	0.25	391.5	3.75	13.84	534.34	215.53	1496.19	27.69
Min	0.17	1.65	0.06	0.15	0.14	23	1.38	3.85	86.92	96.85	480	18.64
Max	0.51	4.88	0.19	0.48	0.45	2223	8.42	49.11	2328	398	3346	50.55
h^2	0.5	0.31	0.42	0.27	0.67	0.33	0.08	0.3	0.58	0.44	0.76	0.63
Genotype (P-Value)	***	*	***	**	**	*	ns	*	ns	***	***	***
Stuttgart (2013-2014)												
Mean	0.12	1.42	0.06	0.21	0.1	64.93	0.89	5.34	161.63	140.25	1172.97	11.55
Min	0.05	0.75	0.03	0.1	0.051	4.47	0.33	0.61	51.45	56.69	351	4.62
Max	0.28	2.34	0.099	0.33	0.19	382.37	1.7	18.16	1038	295	2996	24.97
h^2	0.39	0.18	0.18	0.34	0.38	0.04	0.23	0.01	0.12	0.44	0.45	0.26
Genotype (P-Value)	***	ns	ns	**	**	ns	ns	ns	ns	***	***	*
St. Joseph (2012-2013)												
Mean	0.39	3.48	0.17	0.48	0.32	195.96	1.81	6.62	375.7	28.02	195.97	22.98
Min	0.24	2.71	0.11	0.3	0.18	74.5	0.46	3.55	157.13	15.82	78.3	13.46
Max	0.61	4.87	0.26	0.69	0.48	552.86	9.11	16.77	1010.41	123.41	605.32	227
h^2	0.71	0.67	0.53	0.58	0.66	0.06	0.05	0.27	0.1	0.44	0.37	0.32
Genotype (P-Value)	***	***	***	***	***	ns	ns	*	ns	***	**	*
St. Joseph (2013-2014)												
Mean	0.33	2.94	0.11	0.29	0.18	33.21	2.3	6.71	93.68	13.37	104.1	10.57
Min	0.19	2.25	0.09	0.21	0.12	5.61	1.13	2.16	38.22	5.96	45.22	6.84
Max	0.5	3.79	0.15	0.45	0.31	144.78	7.9	16.03	306.1	28.21	308.42	22.1

[+]Narrow sense heritability estimates for adjusted means; calculated as

$$h^2 = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_E}{r}}$$

***Significant at $p < 0.0001$ level

**Significant at $p < 0.001$ level

*Significant at $p < 0.05$ level

Table 2 Descriptive statistics and analysis of variance of macro and micronutrient concentrations and yield components for 240 wheat genotypes grown under field waterlogging conditions in Stuttgart AR and Saint Joseph LA, 2012-2014

	Macronutrients					Micronutrients						
Source	**Ca**	**K**	**Mg**	**P**	**S**	**Al**	**B**	**Cu**	**Fe**	**Mn**	**Na**	**Zn**
	%	**%**	**%**	**%**	**%**	**ppm**	**ppm**	**ppm**	**ppm**	**ppm**	**ppm**	**ppm**
Stuttgart (2012-14)												
Mean	0.22	2.17	0.08	0.25	0.17	228.26	2.32	9.59	347.99	177.89	1334.33	19.62
Min	0.17	1.7	0.05	0.17	0.14	70.76	1.62	4.41	177.69	109.41	696.47	14.53
Max	0.32	2.76	0.11	0.34	0.23	759.94	3.74	15.43	1198.89	279.99	2836.85	30.81
h2	0.52	0.14	0.5	0.49	0.26	0.09	0.02	0	0.13	0.46	0.59	0.53
Geno (P-value)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.2211	0.0349	0.0001	0.0001	0.0001	0.0001
Geno*year (P-value)	0.0026	0.0023	0.0057	0.3703	0.0008	0.0012	0.2574	0.0142	0.0001	0.039	0.0001	0.0001
St. Joseph (2012-14)												
Mean	0.37	3.3	0.15	0.42	0.27	143.93	1.98	6.66	281.69	23.13	165.35	18.82
Min	0.23	2.82	0.11	0.31	0.21	75.94	1.11	3.76	162.16	14.36	91.38	13.25
Max	0.49	4.01	0.2	0.53	0.36	291.95	4.11	12.29	540.34	56.5	380.56	26.38
h2	0.66	0.35	0.46	0.27	0.35	0.16	0.19	0.02	0.22	0.54	0.53	0.29
Geno (P-value)	0.0001	0.0001	0.1126	0.0006	0.0001	1	0.0001	0.0001	1	0.3774	0.517	0.9023
Geno*year (P-value)	0.0503	0.0001	0.993	0.0947	0.0334	1	0.0029	0.0001	1	1	1	0.9994
All site-years (2012-2014)												
Mean	0.29	2.65	0.11	0.32	0.22	192.11	2.17	8.33	319.57	111.57	833.34	19.28
Min	0.2	2.27	0.09	0.25	0.18	94.31	1.59	4.84	191.02	72.19	450.32	15.66
Max	0.37	3.07	0.14	0.41	0.27	498.35	3.32	12.02	800.67	170.03	1748.37	26.22
h2	0.73	0.20	0.45	0.24	0.18	0.14	0.43	0.28	0.20	0.48	0.57	0.54
Geno (P-value)	0.0001	0.0001	0.0001	0.0001	0.0001	0.758	0.0001	0.0005	0.0001	0.0088	0.0001	0.0001
Geno*loc (P-value)	0.0001	0.0001	0.0001	0.0001	0.0001	0.515	0.2156	0.0501	0.0001	0.4912	0.0001	0.0008

[+]Narrow sense heritability estimates for adjusted means; calculated as

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}$$

Table 3. Descriptive statistics and analysis of variance of yield components for 240 wheat genotypes grown under field waterlogging conditions in Stuttgart AR, growing seasons 2012-2013 and 2014-2015.

Source	Yield Components		
	TKW (g)	**KWS (g)**	**KNPS (No.)**
Stuttgart (2012-2013)			
Mean	33.97	1.04	31.99
Min	20.17	0.29	12.86
Max	48.68	1.84	63.28
H^2	0.72	0.06	0.07
Genotype (P-Value)	0.002***	0.07	0.003
Stuttgart (2014-2015)			
Mean	28.76	0.66	23.06
Min	16.4	0.23	9.99
Max	61.6	1.21	43.52
H^2	0.02	0.8	0.62
Genotype (P-Value)	0.18	0.05***	0.13***
Stuttgart (2012-2015)			
Mean	32.06	0.87	27.63
Min	23.65	0.5	14.91
Max	42.6	1.3	41.73
H^2	0.83	0.32	0.2
Geno (P-value)	0.0001	0.0001	0.0001
Geno*year (P-value)	0.0001	0.0015	0.0082

[+]Narrow sense heritability estimates for adjusted means; calculated as

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}$$

***Significant at $p < 0.0001$ level
**Significant at $p < 0.001$ level
*Significant at $p < 0.05$ level

Table 4. Comparison of genomic selection prediction accuracies and (g_s) phenotypic prediction ($H^{1/2}$) for individual and combined site-year datasets for twelve micro and macronutrients for 240 wheat genotypes grown under field waterlogging conditions in Stuttgart AR and St. Joseph LA, 2012-2014**.**

		Prediction accuracies											
		Macronutrients					**Micronutrients**						
Site-year	**Accuracy**	**Ca**	**K**	**Mg**	**P**	**S**	**Al**	**B**	**Cu**	**Fe**	**Mn**	**Na**	**Zn**
Stuttgart													
2012-2013 (ST13)	r_{gs}	0.29	0.25	0.26	0.20	0.17	0.21	0.14	0.11	0.23	0.30	0.31	0.24
	$H^{1/2}$	0.71	0.56	0.65	0.52	0.82	0.57	0.28	0.55	0.76	0.67	0.87	0.79
	$r_{gs}/H^{1/2}$	0.41	0.45	0.40	0.38	0.21	0.37	0.50	0.20	0.30	0.45	0.36	0.30
2013-2014 (ST14)	r_{gs}	0.22	0.16	0.28	0.26	0.24	-0.10	0.04	-0.02	-0.02	0.26	0.36	0.18
	$H^{1/2}$	0.62	0.43	0.42	0.58	0.62	0.19	0.47	0.01	0.34	0.66	0.67	0.51
	$r_{gs}/H^{1/2}$	0.36	0.36	0.67	0.45	0.39	-	0.08	-	-	0.39	0.54	0.35
Both years (STCOMB)	r_{gs}	0.38	0.30	0.25	0.30	0.22	0.15	0.07	0	0.07	0.36	0.50	0.37
	$H^{1/2}$	0.72	0.37	0.71	0.70	0.51	0.30	0.14	0	0.36	0.68	0.77	0.73
	$r_{gs}/H^{1/2}$	0.53	0.82	0.35	0.42	0.43	0.49	0.53	0	0.20	0.52	0.65	0.51
St. Joseph													
2012-2013 (SJ13)	r_{gs}	0.45	0.39	0.33	0.25	0.41	0.13	0.27	0.18	0.13	0.38	0.27	0.24
	$H^{1/2}$	0.84	0.82	0.73	0.76	0.81	0.24	0.21	0.52	0.32	0.67	0.61	0.56
	$r_{gs}/H^{1/2}$	0.53	0.48	0.46	0.34	0.5	0.53	1.28	0.34	0.41	0.56	0.45	0.42
2013-2014 (SJ14)	r_{gs}	0.21	0.37	0.16	0.30	0.42	-0.06	0.37	-0.02	0.07	0.22	0.37	0.40
Both years (SJCOMB)	r_{gs}	0.43	0.41	0.30	0.22	0.48	0.09	0.42	0.05	0.15	0.39	0.32	0.30
	$H^{1/2}$	0.81	0.59	0.68	0.52	0.59	0.4	0.43	0.14	0.47	0.74	0.73	0.54
	$r_{gs}/H^{1/2}$	0.53	0.70	0.44	0.43	0.81	0.23	0.98	0.36	0.31	0.53	0.44	0.56
Combined analysis													
All site-years combined (STSJALL)	r_{gs}	0.46	0.44	0.35	0.30	0.52	0.15	0.23	0.09	0.06	0.38	0.50	0.39
	$H^{1/2}$	0.85	0.45	0.67	0.49	0.42	0.37	0.65	0.53	0.45	0.69	0.75	0.73
	$r_{gs}/H^{1/2}$	0.54	0.97	0.53	0.61	1.25	0.40	0.36	0.17	0.13	0.55	0.67	0.54

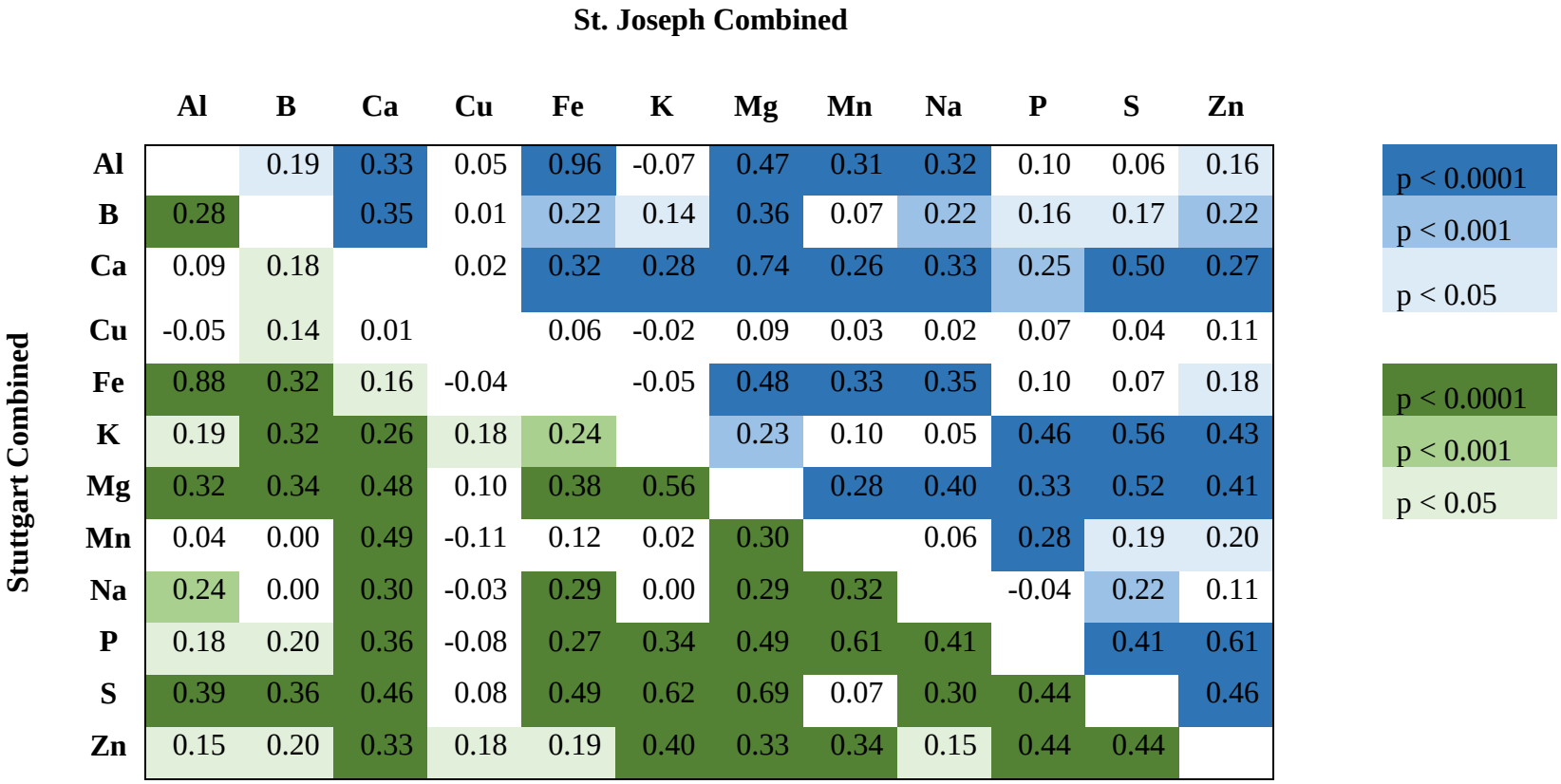

Figure 1. Phenotypic correlations of shoot micro and macronutrient concentrations in 240 winter wheat genotypes grown under waterlogging. Combined analysis for Stuttgart and St. Joseph environments, growing seasons 2012-2013 and 2013-2014. The degree of phenotypic correlations are indicated by green and blue color scale intensities.

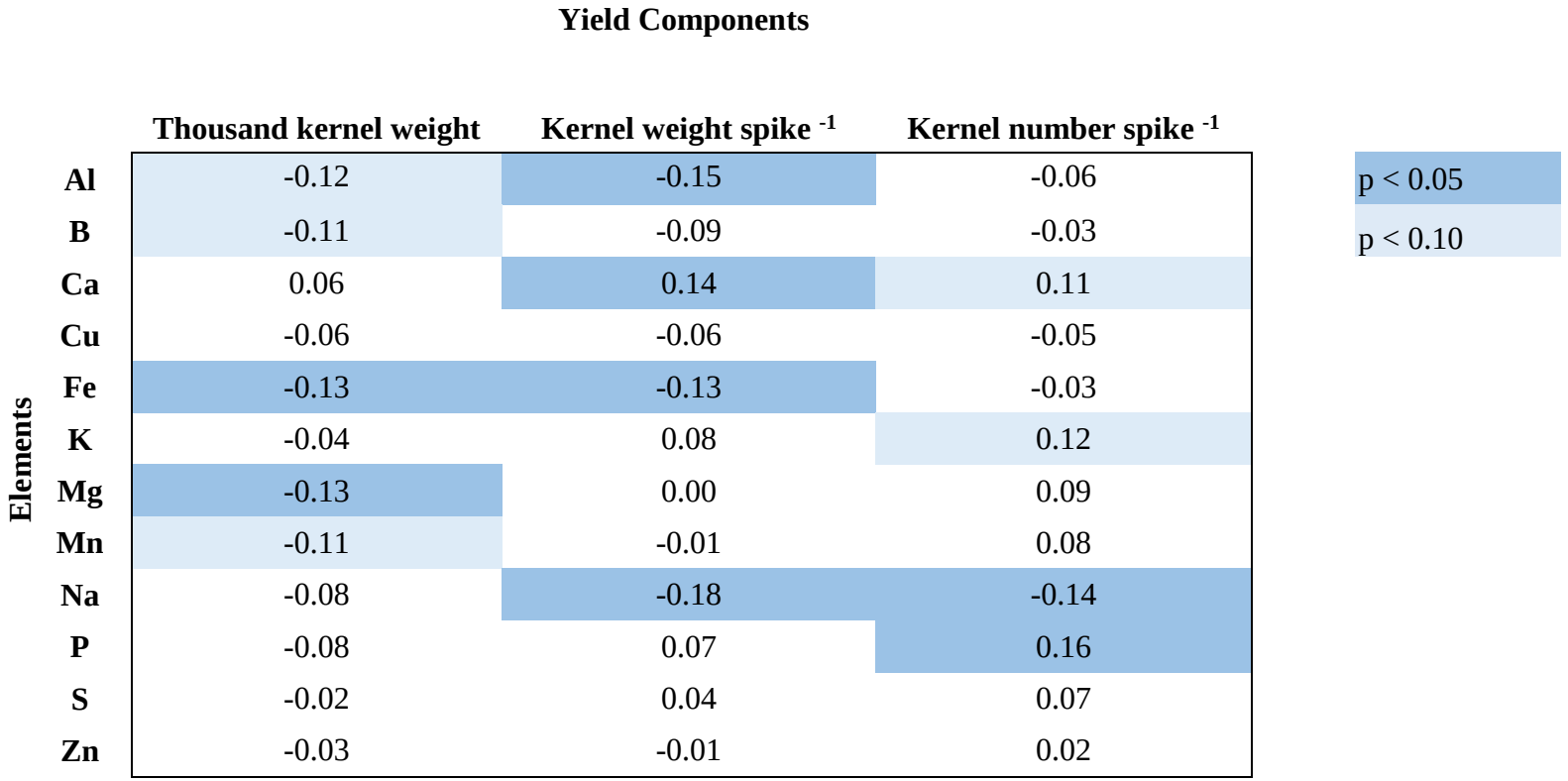

Figure 2. Phenotypic correlations of shoot micro and macronutrient concentrations with yield component traits in 240 winter wheat genotypes grown under waterlogging conditions at the RREC in Stuttgart Arkansas from 2012 to 2014.

References

Ågren GI, Weih M (2012) Plant stoichiometry at different scales: element concentration patterns reflect environment more than genotype. New Phytologist 194:944-952

Arguello MN, Mason RE, Roberts TL, Subramanian N, Acuña A, Addison CK, Lozada DN, Miller RG, Gbur E (2016) Performance of soft red winter wheat subjected to field soil waterlogging: Grain yield and yield components. Field Crops Research 194:57-64

Armstrong W (1979) Aeration in higher plants. Academic Press

Arruda MP, Brown PJ, Lipka AE, Krill AM, Thurber C, Kolb FL (2015) Genomic selection for predicting fusarium head blight resistance in a wheat breeding program. The Plant Genome 8

Boru G, van Ginkel M, Kronstad WE, Boersma L (2001) Expression and inheritance of tolerance to waterlogging stress in wheat. Euphytica 117:91-98

Boyer JS (1982) Plant productivity and environment. Science 218:443-448

Cai SB, Bai GH, Zhang DD (2008) Quantitative trait loci for aluminum resistance in Chinese wheat landrace FSW. Theor Appl Genet 117:49-56

Cambell CR, Plank CO (1992) Organic matter destruction-wet ashing. In: Plank CO (ed) Plant analysis reference procedures for the southern region of the United States, #368 edn. The University of Georgia, Crop & Soild Science Department, p 68

Chaubey CN, Senadhira D (1994) Conventional plant breeding for tolerance to problem soils. In: Yeo A, Flowers T (eds) Soil mineral stresses. Springer Berlin Heidelberg, pp 11-36

Colmer TD, Voesenek LACJ (2009) Flooding tolerance: suites of plant traits in variable environments. Functional Plant Biology 36:665-681

Combs E, Bernardo R (2013) Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. The Plant Genome 6

Crossa J, Campos Gdl, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genetics 186:713-724

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One 3:e3395

Donohue SJ, Aho. DW (1992) Determination of P, K, Ca, Mg, Mn, Fe, Al, B, Cu, and Zn in plant tissue by inductively coupled plasma (ICP) emission spectroscopy. C.O. Plank (ed.). Plant Analysis Reference Procedures for the Southern Region of the United States. Georgia Agric. Expt. Sta., Southern Coop. Ser. Bul. 368.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with r package rrBLUP. The Plant Genome 4:250-255

Foy CD (1984) Physiological effects of hydrogen, aluminum, and manganese toxicities in acid soil. Soil acidity and liming:57-97

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346

Gotoh S, Patrick WH (1974) Transformation of iron in a waterlogged soil as influenced by redox potential and pH. Soil Science Society of America Journal 38:66-71

Hayes BJ, Bowman PJ, Chamberlain AC, Verbyla K, Goddard ME (2009) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution 41:51

Heffner EL, Jannink J-L, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Science 51:2597-2606

Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic selection for crop improvement. Crop Sci 49

Hoffstetter A, Cabrera A, Huang M, Sneller C (2016) Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. G3 (Bethesda, Md) 6:2919-2928

Houtz RL, Nable RO, Cheniae GM (1988) Evidence for effects on the in vivo activity of ribulose-bisphosphate carboxylase/oxygenase during development of Mn toxicity in tobacco. Plant physiology 86:1143-1149

Jiang D, Fan XM, Dai TB, Cao WX (2008) Nitrogen fertiliser rate and post-anthesis waterlogging effects on carbohydrate and nitrogen dynamics in wheat. Plant Soil 304:301-314

Kahlown MA, Azam M (2002) Individual and combined effect of waterlogging and salinity on crop yields in the Indus basin. Irrig Drain 51:329-338

Kerridge P (1968) Aluminum toxicity in wheat (*Triticum aestivum* Vill., Host). Oregon State University. Ph.D. Diss., Microfilms..69–4086. Ann Arbor, Michigan (Diss. Abstr. 29B.-3159).

Khabaz-Saberi H, Barker SJ, Rengel Z (2012) Tolerance to ion toxicities enhances wheat (*Triticum aestivum* L.) grain yield in waterlogged acidic soils. Plant Soil 354

Khabaz-Saberi H, Rengel Z (2010) Aluminum, manganese, and iron tolerance improves performance of wheat genotypes in waterlogged acidic soils. Journal of Plant Nutrition and Soil Science 173:461-468

Khabaz-Saberi H, Setter TL, Waters I (2005) Waterlogging induces high to toxic concentrations of iron, aluminum, and manganese in wheat varieties on acidic soil. Journal of Plant Nutrition 29:899-911

Kristensen PS, Jahoor A, Andersen JR, Cericola F, Orabi J, Janss LL, Jensen J (2018) Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. Frontiers in Plant Science 9:69. doi: 10.3389/fpls.2018.00069

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14):1754-60. doi: 10.1093/bioinformatics/btp324.

Lopez-Cruz M, Crossa J, Bonnett D, Dreisigacker S, Poland J, Jannink J-L, Singh RP, Autrique E, de los Campos G (2015) Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. G3: Genes|Genomes|Genetics

Lozada DN, Mason RE, Babar MA, Carver BF, Guedira G-B, Merrill K, Arguello MN, Acuna A, Vieira L, Holder A, Addison C, Moon DE, Miller RG, Dreisigacker S (2017) Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. Euphytica 213:222

Ma JF, Chen ZC, Shen RF (2014) Molecular mechanisms of Al tolerance in gramineous plants. Plant and Soil 381:1-12

Manickavelu A, Hattori T, Yamaoka S, Yoshimura K, Kondou Y, Onogi A, Matsui M, Iwata H, Ban T (2017) Genetic Nature of Elemental Contents in Wheat Grains and Its Genomic Prediction: Toward the Effective Use of Wheat Landraces from Afghanistan. PLoS One 12 (1): e0169416.

Meuwissen T.H.E, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819-1829

Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S (2015) LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3 (Bethesda, Md) 5:2383-2390

Munch JC, Hillebrand T, Ottow JCG (1978) Transformations in the Feo/Fed ratio of pedogenic iron oxides affected by iron-reducing bacteria. Canadian Journal of Soil Science 58:475-486

Okeke UG, Akdemir D, Rabbi I, Kulakow P, Jannink JL (2017) Accuracies of univariate and multivariate genomic prediction models in African cassava. Genetics, Selection, Evolution : GSE 49:88

Plank C.O., Donohue. SJ (2000) Small Grain: Barley, Oats, Rye, Wheat. In: Plank CO (ed) Reference sufficiency ranges for plant analysis in the Southern Region of the Unites States -Field Crops-. North Carolina Department of Agriculture and Consumer Services Agronomic Division

Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M, Jannink J-L (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. The Plant Genome 5:103-113

Ponnamperuma FN (1972) The chemistry of subemerged soils. Advances in agronomy 24 29-96.

Resende MDV, Resende MFR, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ, Kilian A, Grattapaglia D (2012) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist 194:116-128

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179:161-173

Saint C, Burgueño J, Crossa J, Fuentes Dávila G, Figueroa López P, Solís Moya E, Ireta Moreno J, M Hernández Muela V, M Zamora Villa V, Vikram P, Mathews K, Sansaloni C, Sehgal D, Jarquin D, Wenzl P, Singh S (2016) Genomic prediction models for grain yield of spring bread wheat in diverse agro-ecological zones. 6:27312

Sairam RK, Kumutha D, Ezhilmathi K, Deshmukh PS, Srivastava GC (2008) Physiology and biochemistry of waterlogging tolerance in plants. Biol Plant 52:401-412

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451-461

Setter TL, Waters I (2003) Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. Plant and Soil 253:1-34

Setter TL, Waters I, Sharma SK, Singh KN, Kulshreshtha N, Yaduvanshi NP, Ram PC, Singh BN, Rane J, McDonald G, Khabaz-Saberi H, Biddulph TB, Wilson R, Barclay I, McLean R, Cakir M (2009) Review of wheat improvement for waterlogging tolerance in Australia and India: the importance of anaerobiosis and element toxicities associated with different soils. Ann Bot 103:221-235

Shabala S (2011) Physiological and cellular aspects of phytotoxicity tolerance in plants: the role of membrane transporters and implications for crop breeding for waterlogging tolerance. New Phytologist 190 (2):289-98.

Socha AL, Guerinot ML (2014) Mn-euvering manganese: the role of transporter gene family members in manganese uptake and mobilization in plants. Frontiers in Plant Science 5:106

Steffens D, Hütsch BW, Eschholz T, Lošák T, Schubert S (2005) Water logging may inhibit plant growth primarily by nutrient deficiency rather than nutrient toxicity. Plant, Soil and Environment vol. 51:545–552

Stevenson FJ, Cole MA (1999) Cycles of Soil: Carbon, Nitrogen, Phosphorus, Sulfur, Micronutrients 2ndEdition. Wiley-Interscience. NewYork

Thavamanikumar S, Dolferus R, Thumma BR (2015) Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. G3-Genes Genomes Genetics 5:1991-1998

Trought MCT, Drew MC (1980) The development of waterlogging damage in wheat seedlings (*Triticum aestivum* L.). Plant and Soil 54:77-94

Velu G, Crossa J, Singh RP, Hao Y, Dreisigacker S, Perez-Rodriguez P, Joshi AK, Chatrath R, Gupta V, Balasubramaniam A, Tiwari C, Mishra VK, Sohu VS, Mavi GS (2016) Genomic prediction for grain zinc and iron concentrations in spring wheat. Theoretical and Applied Genetics 129:1595-1605

Wang W, Zhao XQ, Hu ZM, Shao JF, Che J, Chen RF, Dong XY, Shen RF (2015) Aluminium alleviates manganese toxicity to rice by decreasing root symplastic Mn uptake and reducing availability to shoots of Mn stored in roots. Ann Bot 116:237-246

Westerman RL (1987) Soil reaction: Acidity, alkalinity and salinity. In: Heyne EG (ed) Wheat and wheat improvement 2nd ed Agron Mono. ASA, CSSA, SSSA, Madison, WI, pp 340-344

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. Theoretical and Applied Genetics 116:815-824

Yaduvanshi NPS, Setter TL, Sharma SK, Singh KN, Kulshreshtha N (2012) Influence of waterlogging on yield of wheat (*Triticum aestivum*), redox potentials, and concentrations of microelements in different soils in India and Australia. Soil Research 50:489

Ziyomo C, Bernardo R (2013) Drought tolerance in maize: indirect selection through secondary traits versus genomewide selection. Crop Science 53:1269-1275

CHAPTER III

GENOME WIDE ASSOCIATION ANALYSIS OF MICRO AND MACRONUTRIENT CONCENTRATIONS UNDER WATERLOGGING STRESS IN WHEAT (*TRITICUM AESTIVUM* L.)

Abstract

Waterlogging impacts up to 15 million hectares of wheat globally and tolerance to micro and macronutrient toxicities and deficiencies play a key role in overcoming this constraint. The objective of this study was to identify marker-trait associations (MTA) for accumulation of nutrient concentrations under field waterlogging stress using a genome wide association study (GWAS). A panel of 240 soft red winter wheat lines was subjected to soil waterlogging over two seasons at the Rice Research and Extension Center in Stuttgart, AR and the Northeast Louisiana Research Station in St. Joseph, LA, US. Accumulation of macronutrients calcium (Ca), potassium (K), magnesium (Mg), phosphorus (P) and sulfur (S) and micronutrients aluminum (Al), boron (B), copper (Cu), iron (Fe), manganese (Mn), sodium (Na) and zinc (Zn) were determined in wheat shoots post-waterlogging stress using inductively coupled plasma atomic emission spectroscopy. Principal component analysis showed consistent grouping of micro and macronutrient concentrations and an [Al] and [Fe] clustering present across all four site-years. In total, 78 MTA were identified using a mixed model GWAS approach that included both kinship and the *Q* matrix to account for population structure and genetic relatedness. Forty-seven highly significant ($p < 0.00001$), twenty-three very significant and consistent ($p < 0.0005$) and eight significant and consistent ($p < 0.001$) MTA were identified. These markers spanned all wheat chromosomes with the exception of 2D, 3D, 4A, 4B, 5D, 6A and 7A. Blastx analysis identified 18 MTA to be associated either with nucleic acid binding, RNA - DNA hybrid ribonuclease activity and zinc ion binding.

Introduction

Soil waterlogging (WL) and flooding resulting from high rainfall and poor soil drainage are major limitations to crop production both globally (Setter and Waters 2003) and in the US (Bailey-Serres et al. 2012). Waterlogging results in low soil 0_2 concentration and anoxia, with subsequent hypoxia in plant roots (Ahmed et al. 2013; Setter et al. 2009). Under anoxic soil conditions, a reduced redox potential results in increased concentrations of soil NO^{-2}, Mn^{2+}, Fe^{2+} and increased plant uptake, potentially leading to elemental toxicities (Ponnamperuma 1972). Plants cope with increased availability of micro and macronutrients through either exclusion or tissue tolerance mechanisms. Exclusion mechanisms are characterized for some elements, including aluminum, in which organic compounds are released at the root tip to chelate aluminum ions, forming nontoxic compounds (Kochian et al. 2015; Yang et al. 2008). Other avoidance mechanisms include cell wall modification to impede toxic elements from entering the root tissue or immobilization in the roots (Wang et al. 2004). Tissue tolerance mechanisms include sequestration and storage in vacuoles through binding with peptides, proteins or organic compounds (Garbisu and Alkorta 2001; Harborne 1989).

Association mapping (AM) is used to identify significant associations between molecular markers and phenotypic traits in a panel of lines (Gupta et al. 2005) that can be exploited for marker assisted selection (Bentley et al. 2014; Hamblin et al. 2011; Zhu et al. 2008). The basis of AM is linkage disequilibrium (LD) in which alleles in a chromosome located at different loci are non-randomly associated (Flint-Garcia et al. 2003; Mackay and Powell 2007). AM captures the historical recombination events present within a germplasm panel and identifies the genetic control of quantitative traits at potentially higher resolution compared to a bi-parental approach (Huang and Han 2014; Myles et al. 2009).

Much of the research related to the genetic control of exclusion mechanisms in wheat has focused on aluminum uptake, particularly in low pH soils, and to a lesser extent on boron and sodium. Tolerance of wheat to Al^{3+} has been associated with the malate transporter *Alt1* locus located on chromosome 4DL, which results in improved root growth (Delhaize et al. 1993; Ma et al. 2005). Similar genetic studies on Al uptake have been performed using both bi-parental (Dai et al. 2013) and association mapping (Froese and Carter 2016; Navakode et al. 2014; Raman et al. 2010) approaches, with QTL identified on at least 16 of the 21 wheat chromosomes. Tolerance to high levels of boron, particularly in irrigated systems, is conditioned by the *Bo1* locus, which was mapped to a 1.8 cM region on chromosome 7BL (Schnurbusch et al. 2007). Oyiga et al. (2017) used GWAS to detect MTA for exclusion of potassium and sodium in a panel of 150 winter wheat lines and identified 187 significant ($P = 8.22\ E^{-05}$ to $5.10\ E^{-06}$). This included two candidate genes, a zinc II ion transmembrane transporter on chromosome 1BS and a sodium ion transport protein on chromosome 1DL.

For US wheat germplasm, previous studies have focused on micronutrient exclusion mechanisms under dryland or low pH soil conditions (discussed above) with limited knowledge of these mechanisms under soil waterlogging. The southeastern US soft red winter wheat germplasm pool is the most well characterized in terms of tolerance to soil waterlogging (Arguello et al. 2016; Ballesteros et al. 2015; Collaku and Harrison 2002, 2005).
(Arguello et al. 2016; Ballesteros et al. 2015; Collaku and Harrison 2002, 2005). However, with the exception of the study by Ballesteros et al. (2015) little is known about the mechanism of the observed tolerance and there are no reports of the genetic control of micro and macronutrient exclusion under waterlogging. It was hypothesized that the variable response to waterlogging in terms of yield production as reported by Arguello et al. (2016) is due in part to micro and

macronutrient exclusion mechanisms. Hence, the purpose of this study was to evaluate a set of 240 SRWW lines under field waterlogging conditions and use GWAS to identify marker-trait associations and potential candidate genes.

Materials and methods

Experimental design and trait measurement

The association panel (AP) used for this study consisted of 240 inbred lines of soft red winter wheat. Information on the population and collection and analysis of the phenotypic data was previously described in Chapter 2 of this dissertation. Briefly, experiments were conducted at the Rice Research and Extension Center in Stuttgart (ST) Arkansas, AR and the Northeast Louisiana Research Station in St. Joseph (SJ), Louisiana during the growing season of 2012-2013 (13) and 2013-2014 (14). The population was sown in a randomized complete block design (RCBD) with two replications in ST13, SJ13 and SJ14 and in an augmented incomplete block design with two repeated check lines ('Jamestown' and 'Pioneer Brand 26R20') in ST14. A WL treatment was imposed by establishing 0.30 m high levees surrounding the experimental field beginning at Feekes 4 and continuing for a duration of 14 to 28 days depending on the site-year. A 0.10 m^2 section per plot was harvested 24 hours after WL treatment and processed according to Donohue and Aho. (1992) for determination of total concentrations of macronutrients calcium (Ca), potassium (K), magnesium (Mg), phosphorus (P) and sulfur (S) and micronutrients aluminum (Al), boron (B), copper (Cu), iron (Fe), manganese (Mn), sodium (Na) and zinc (Zn) using inductively coupled plasma atomic emission spectroscopy (ICP-AES). Soil samples from ST and SJ sites were evaluated using the Spectro Arcos IPC-OES analyzer were pH, electric conductivity (EC) and micro and macronutrient concentrations of B, Ca, Cu, Fe, K, P, Mn, Mg, S and Zn were determined (Supplementary 1).

Genotyping and SNP calling

Genome wide marker data for the AP panel was generated using genotype by sequencing (GBS) with SNP calling using the TASSEL 5 GBSv2 pipeline (Glaubitz et al. 2014) as previously described in Chapter 2 of this dissertation. Briefly, raw SNP data was filtered to remove taxa with more than 90% missing data and then for bi-allelic SNPs with minor allele frequency (MAF) ≥ 5%, missing data ≤ 50% and heterozygosity ≤ 10%. After filtering, LD-kNNi in TASSEL was used for imputation of missing data (Money et al. 2015) resulting in a total of 62,375 SNP markers for analysis.

Linkage disequilibrium and principal component analysis

Linkage disequilibrium among markers was calculated using the sliding window option in TASSEL 5.2.31 (Bradbury et al. 2007). Only SNPs with MAF > 5% and with defined genetic positions were used in the analysis with 43,072 markers meeting these criteria. LD was calculated for each chromosome (Table 1) and for the three (A, B and D) wheat genomes. Pair wise LD was calculated using the square allele frequency correlations r^2 and the number of significant pairs ($p < 0.01$) was determined using 1,000 permutations.

To determine association between the traits, a principal component analysis (PCA) was performed using the PRINCOMP function in R Studio. Contribution of each variable to the first two principal components (PC) were illustrated using bi-plots.

Population structure

Population structure of the AP was evaluated using fastSTRUCTURE (Raj et al. 2014) and the algorithm was run on 92,916 markers. Model complexity was determined running multiple options of K =1 to 10 with 20 runs per K evaluated. Cluster identification was performed with the deviance residuals and variation parameters using the simple prior tool integrated in the

program (Raj et al. 2014). Based on the highest value for coefficient of membership *(Q)* lines were grouped in three subpopulations: 1) Two hundred and thirty three lines assigned to subpopulation *Q1*, 2) four lines assigned to subpopulation *Q2*, and 3) three lines assigned to subpopulation *Q3*. Bar plot for Q coefficients were visualized using web application POPHELPER (Francis 2017).

Genome wide association analysis

Best linear unbiased estimators (BLUEs) for each trait and line were estimated within individual site-years and per location for GWAS analysis as described previously in Chapter 2 of this dissertation. Association mapping analyses were performed using the compressed mixed linear model (CMLM) implemented in the genome association and prediction integrated tool (GAPIT) package in R (Lipka et al. 2012). Three analyses were conducted evaluating different mixed model approaches: 1) a *K* only approach with a kinship matrix (K) for line relatedness generated from Tassel 5 was used in the model, 2) a *K* + *PC* model, in which the kinship matrix and the first three principal components were used, and 3) a *K* + *Q* model in which the kinship matrix and the Q matrix generated from fastSTRUCTURE were used in order to control for line relatedness and population structure respectively (Yu et al. 2006). The following is the model *K* + *Q* used to account for genetic relatedness among the association panel

$$y = \mu + x\beta + Qv + Zu + e$$

where y is the vector of observed phenotype; μ is the mean; x is the effect of the SNP; v is a vector of population effects, u is the random effects due to genetic relatedness. Variances of random effects are assume to be Var ($\boldsymbol{u}$) = $2KV_g$ and Var (e) = RV_e; where $\boldsymbol{Z}$ is the kinship matrix across all genotypes, e is the vector of residual effects; $\boldsymbol{Q}$ is a matrix from fastSTRUCTURE relating y to v. Vg is the genetic variance; and Ve is the residual variance (Yu et al. 2006). Complementary to

the GWAS analysis marker positions (bp) were used to determine the genomic regions of the MTA identified. The IWGSC RefSeq v1.0 browser and the Blatx tool aligned Triticum aestivum dataset (TGACv1) available in *EnsemblPlants* (https://plants.ensembl.org/index.html) were used to investigate the molecular and biological function of the MTA identified.

Results

Association between traits using principal components analyses

Principal component analysis was generally consistent for all traits across site years with a similar grouping among micro and macronutrients observed and a clustering of Fe and Al in all environments. A negative correlation among PC1 and all twelve element concentrations was observed in all four environments. In addition, PC1 explained a similar percent of the phenotypic variance across site-years with 32.2, 32.21, 33.7 and 35.50% for ST13, ST14, SJ13 and SJ14 respectively (Fig 1). Likewise, PC2 had similar values in ST13 (15.3%), ST14 (18.0%), SJ13 (15.35%) and SJ14 (18.74%). Micronutrients Al and Fe were positively correlated with PC2 in all environments with the exception of SJ14 in which a negative correlation was observed. Clustering of K, Zn, P and S was consistent across both years in SJ (Fig. 1c and 1d). Clustering of macronutrients was observed in Stuttgart, with Mg, S and K in ST13 (Fig. 1a) and Ca, Mg, P, and S in ST14 (Fig. 1b).

Linkage disequilibrium and population structure

Linkage disequilibrium at the whole genome level identified 2,081,476 loci pairs, including 1,631,419 (11.07%) intrachromosomal pairs in significant LD ($p > 0.005$) and 130,222 (6.42%) pairs in complete LD ($R^2 = 1$) (Tassel 5.3) (Table 1). A negative correlation between linkage disequilibrium and physical distance was observed ($r = -0.37$) as well as a negative correlation among linkage disequilibrium and P-values. Analysis of genetic structure of the 240 lines

determined the likely number of clusters at *K=3* (Fig. 2), but with only seven lines grouped in clusters *Q2* and *Q3***.** Given the small number, these outlier lines were removed from further analysis. Principal component analysis identified two subpopulations, which aligned with the presence of the *Sr36* gene translocation in 63 of the 240 lines. Similar results were previously reported by Benson et al. (2012) in which this introgression was observed in 57 of 251 wheat lines and commonly present in lines originating from southern US latitudes.

Marker trait Associations

Three models were used to identify marker trait associations (MTA) between 62,376 GBS SNP markers and the concentration of micro and macronutrients under WL stress. Although non-significant variation occurred across models, the observed P values and cumulative distribution of P for the most highly heritable trait determined the *K-Q* approach as the best method to account for population structure and genetic relatedness among individuals (Bordes et al. 2014). Moreover, results from the Q-Q plots showed sharp deviations in the tail area of the *P* value distributions, indicating this model controlled for false-positives and false negative associations (Kaler et al. 2017).

Summary of MTA for microelement concentrations

Three different P values were used as a threshold to declare MTA: 1) highly significant ($p < 0.00001$), 2) very significant and consistent ($p < 0.0005$ and present in at least two site-years) and 3) significant and consistent ($p < 0.001$ and present in at least two site years) (Manickavelu et al. 2017). In total, 78 MTA were identified across the A (19), B (46) and D (9) genomes, with four significant markers unassigned to a chromosome (Supplementary 2). MTA were identified across all wheat chromosomes with the exception of 2D, 3D, 4A, 4B, 5D, 6A and 7A.

Highly significant MTA

Forty seven highly significant ($p < 0.00001$) MTA were identified, including for Al, B, Cu, Fe, Na, P, S and Zn across chromosomes 1A, 1B, 1D, 2A, 2B, 3A, 3B, 4D, 6B, 7B, 7D. Positive effects (higher concentrations) were observed for all MTA with respect to the minor allele. Several MTA were localized in the same genomic region or within a distance less than 10 million base pairs (bp) and having identical or highly similar MAF and allelic effect values (Supplementary 2). Therefore, only markers with the most significant P values within a cluster of MTA are reported in Table 2. For example, [Na] had the highest number of MTA with 28 markers identified on chromosomes 1A, 1B and 2A which explained from 15 to 23% of the phenotypic variation. Twenty four of these markers were localized within a genomic distance of 3 million bp. Allelic effects for [Na] ranged from 16.86 to 315.79 ppm with markers *S2A_612152711* and *S2A_612152741* having the largest positive allelic effect and markers *S1B_599416338* and *S1B_599759697* the most highly significant MTA identified in this study ($p < 1.5E-06$).

Six MTA for [Fe] were identified in chromosomes 1A, 1B, 1D, 6B and 7D, accounting for 11 to 17% of the phenotypic variance (Fig 3). Marker *S1A_192014751* showed the highest allelic effect (153.44 ppm) for [Fe]. In SJ13, multi trait marker *S7D_610181301* was associated with [Al] and [Fe] with a positive allelic effect of 34.79 and 59.7 ppm, respectively. In total, twenty-eight MTA were identified on chromosome 1B and were associated with Na, Fe and [Zn].

Screening of the MTA genomic regions found a reported molecular function for eighteen of the 47 highly significant MTA. Functions included nucleic acid binding and zinc ion binding as molecular functions for markers *S1B_599652390, S1B_599652390, S1B_599652508,*

S1B_599652549, *S7B_117334234*, and *S7B_117334235* with *S1B_599652508* and *S1B_599652549* which was also an MTA for [Zn] (Table 3).

Very significant and consistent MTA

Twenty three very significant ($p < 0.0005$) and consistent (in at least two of the four site-years) MTA were identified, including for Al, Ca, Fe, Na and Zn and located on chromosomes 1A, 1B, 5B, 6B, 7B and 7D. Sixteen MTA were identified in common between the ST13 and SJ14 site-years. This included nine MTA for [Fe] in chromosome 1A, which explained from 8 to 10% of the phenotypic variance, three markers associated with [Na] and located in chromosomes 1B and 5B, and four MTA for [Zn] located in chromosomes 1B, 7B and 7D.

Three MTA associated with [Ca] were consistent across SJ13 and SJ14 including *S6B_309688673* and *S6B_309688674*. Negative allelic effects (-0.03%) were observed for these markers, which accounted for 23% of the phenotypic variance. Similarly, marker *S7D_579146731* associated with [Na] was consistent across SJ13 and SJ14 with a positive allelic effect of 38.3 ppm. For Stuttgart, a single MTA for [Fe] was identified with allelic effects of (21.15ppm) ST14 and (1.64 ppm) ST13. Three additional MTA for [Al] were identified in both SJ14 and ST14 but could not be assigned to a chromosome position (Table 4).

Significant and consistent MTA

In total, eight MTA were both significant ($p < 0.001$) and consistent across two or more site-years, including for K, Mg, Na, P and Zn. Markers *S6D_466814084* and *S6D_466814117* were associated with [K] in both ST13 and SJ14 and showed a negative allelic effect of -0.11% (Table 4). In addition, 76 MTA ($p < 0.0005$) identified in a single environment but associated with more than one element were identified (Supplementary 3). For example, in SJ13 marker

S2A_694912728 was associated with Al, Fe and Mn with allelic effects of 18.05, 28.87 and 2.34 ppm respectively.

Candidate Genes

Screening of the MTA genomic regions found a reported molecular function for 18 MTA identified in this study. Molecular functions included nucleic acid binding, zinc ion binding, structural constituent of ribosome and RNA-DNA hybrid ribonuclease activity were found to be reported for markers located on chromosomes 1A, 1B, 6B, 7B and associated with Al, B, Fe, Na, P and [Zn]. Reported biological functions of DNA integration and translation processes were reported for ten and one MTA, respectively (Table 3).

Discussion

Determining the genetic basis of waterlogging tolerance is an important step for wheat improvement. GWAS studies are used to determine marker trait associations to identify the underlying genetic effect of quantitative traits (Korte and Farlow 2013). To the best of our knowledge, this is the first study evaluating wheat micro and macronutrient concentrations and their potential role in elemental toxicities under soil WL stress using a GWAS approach. A total of 78 MTA were identified, including 47 highly significant ($p < 0.00001$), 23 very significant and consistent ($p < 0.0005$) and eight significant and consistent ($p < 0.001$) MTA.

In total, 63 MTA identified in this study have positive allelic effects indicating these alleles to have been previously introduced and overtime become present at a higher frequency in this set of breeding lines and cultivars through either natural selection or breeding efforts. On the other hand, 15 MTA showing negative allele effects were also observed, where the minor allele was associated with reduced concentrations. Markers *S6B_309688673* and *S6B_309688674* associated with [Ca] showed consistent negative allelic effects values (-0.03) in two site-years.

These markers are likely to be located in the same genomic region and with a MAF of 0.07 could potentially be used as a new source of alleles for modifying [Ca]. On the other hand, nine MTA for [Fe] identified in the same genomic region on chromosome 1A, showed positive allelic effect in SJ14 and negative effect in ST13. Further research is necessary to determine the value of this MTA for reducing Fe uptake during soil waterlogging stress due to its site-specific effect.

Comparison of MTA to previous reported QTL

A comparison to the current literature citing the genetic control of waterlogging tolerance in wheat found strong agreement with the results obtained in this study. MTA were in the same chromosomes reported by Ballesteros et al. (2015) who identified ten genomic regions associated with adaptive traits under waterlogging conditions. In addition, both studies identified no MTA localized to chromosomes 2D, 3D, 4A, 4B, 5D, 6A and 7A. SNP markers *S6B_309688673* and *S6B_309688674* associated with [Ca] and showing a negative allelic effect were located on chromosome 6B in a similar position to a QTL reported by Yu and Chen (2013) with flanking markers *XksuH14* and *Xfbb364* associated with root and shoot dry weight under WL stress. St. Burgos et al. (2001) reported four QTLs associated with flooding tolerance located on wheat chromosomes 2B, 3B and 5A, and similarly in our study, four MTA were identified in the same chromosomes and associated with [Fe], [Mg], [S] and [P]. In addition, markers *S7B_117334234* and *S7B_117334235* were associated with [B] and were potentially in the same region as the *Bo1* QTL, which is reported to control shoot B accumulation. (Jefferies et al. 2000)

Comparison among QTL detected in single studies can be challenging due to environmental variation that results in QTL not being present across environments (Acuña-Galindo et al. 2015). Additionally, physical properties, soil chemistry, duration and severity of the treatment and environmental factors such a microorganism content and soil temperature have

been reported as major factors influencing variability of genotype performance under waterlogging (Drew and Lynch 1980; Kirk et al. 2003; Setter et al. 2008). Clustering among [Fe] and [Al] across all locations was in agreement with previous studies in which an increase of [Fe] and [Al] were observed in wheat lines evaluated in acidic soils (Khabaz-Saberi and Rengel 2010; Khabaz-Saberi et al. 2005; Kochian et al. 2004). However, Foy and Fleming (1982) reported an antagonistic effect in which a reduction in shoot [Fe] was observed due to high [Al]. This could be due to the reasons explained above, in which repeatability and results of tolerant genotypes under waterlogging conditions could be highly variable across environments.

Conclusions

In this study, 78 MTA associated with micro and macronutrient concentrations under waterlogging conditions were identified. These SNP markers were located on the same chromosomes and in some cases chromosome regions of previously identified QTL for waterlogging tolerance. As such, they provide a potential mechanism for future dissection of these QTL regions.

Table 1. Summary of linkage disequilibrium (LD) analyses for intrachromosomal marker pairs using Tassel 5.3 for 240 soft red winter wheat lines.

Genome	Total no of pairs	Mean r^2 for all pairs	No. significant pairs [a]	Significant pairs (%)	Mean r^2 for significant markers	No pairs in complete LD [b]	Pairs in completely LD (%)
Genome A	766,725	0.29	600,915	78.37	0.37	34,822	4.54
Genome B	1,112,225	0.36	894,420	80.42	0.44	78,273	7.04
Genome D	202,526	0.32	136,084	67.19	0.47	17,127	8.46
Genome	2,081,476	0.32	1,631,419	78.38	0.43	130,222	6.26

[a] Significant marker pairs, p value < 0.005

[b] Marker pairs with r^2 value equal to 1.0 were regarded to be in complete LD

Table 2. Highly significant (p < 0.00001) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14) and St. Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

Highly significant < 0.00001						
SNP	**Trait**	**Environment**	**P. value**	**MAF**	**R^{2a}**	**Allelic Effect [b]**
S7D_610181301 (2)	Al / Fe	SJ13	2.5E-06	0.06	0.15	34.79
S7B_117334234 (2)	B	SJ14	1.3E-06	0.06	0.19	0.61
SUN_244556964	B	SJ15	8.0E-06	0.08	0.18	0.45
S3A_702519039	B	SJ13	6.2E-06	0.11	0.14	0.22
S4D_238865033 (2)	Cu	ST13	9.9E-06	0.06	0.16	2.36
S1D_41196155	Fe	ST13	9.9E-07	0.06	0.12	152.54
S1B_46140047 (2)	Fe	ST13	2.0E-06	0.10	0.11	121.74
S6B_111561002	Fe	ST13	2.9E-06	0.08	0.11	131.96
S1A_192014751	Fe	ST13	1.4E-06	0.07	0.11	153.44
S1A_176108383 (2)	Na	SJ13	8.1E-06	0.06	0.15	42.53
S1B_599416338 (24)	Na	ST13/ST14/SJ14	1.5E-08	0.22	0.23	16.86 -183.89
S2A_612152711 (2)	Na	ST13	2.2E-06	0.05	0.19	315.79
S3B_546349156	P	SJ14	7.9E-06	0.14	0.19	0.02
S2B_77204516	S	SJ14	3.7E-06	0.34	0.22	0.01
S7B_568062247	Zn	ST13	9.1E-06	0.11	0.14	2.05
S1A_584433749 (3)	Zn	SJ14	4.3E-06	0.11	0.22	1.27 - 3.66

[a] Reflect the phenotypic variation explained by the marker, R^2 of the model with SNP calculated in GAPIT package in R

[b] Allelic effects with respect to the minor allele

Numbers in parenthesis represent the number of marker in linkage (Less than 10 Million bp)

Table 3. Molecular and biological functions of the significant MTA identified on the 240 lines of the association mapping panel.

SNP	Trait	Molecular function	Biological function	Gene Hit
S1B_597722660	Na	Nuclei acid binding - RNA-DNA hybrid ribonuclease activity	No reported	TRIAE_CS42_1BL_TGACv1_030807_AA0101190
S1B_598061897	Na	Nucleic acid binding - RNA- DNA hybrid ribonuclease activity	No reported	TRIAE_CS42_1BL_TGACv1_030807_AA0101190
S1B_599416338	Na	Structural constituent of ribosome	Translation	TRIAE_CS42_1BS_TGACv1_051115_AA0177850.1
S1B_599652390	Na	Nucleic acid binding / zinc ion binding	DNA integration	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S1B_599652508	Zn	Nucleic acid binding / zinc ion binding	DNA integration	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S1B_599652549	Zn	Nucleic acid binding / zinc ion binding	DNA integration	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S1B_599660677	Na	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S1B_599660679	Na	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S1B_599759697	Na	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_1BL_TGACv1_030539_AA0093500
S6B_111561002	Fe	Nucleic acid binding	DNA integration	TRIAE_CS42_6BL_TGACv1_500432_AA1604470
S7B_117334234	B	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_7BL_TGACv1_582804_AA1918670
S1A_373360656	Fe	Nucleic acid binding	DNA integration	TRIAE_CS42_1AL_TGACv1_000455_AA0012460
S1A_376790586	Fe	Nucleic acid binding	DNA integration	TRIAE_CS42_1AL_TGACv1_000070_AA0002290
S1A_379438389	Fe	Nucleic acid binding / RNA - DNA Hybrid ribonuclease activity	DNA integration	TRIAE_CS42_1AL_TGACv1_000024_AA0000740
S7D_528997111	Zn	Nucleic acid binding / RNA - DNA Hybrid ribonuclease activity	DNA integration	TRIAE_CS42_7DL_TGACv1_603035_AA1974270
SUN_334466963	Al	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_U_TGACv1_669171_AA2155140.1
SUN_334466979	Al	Nucleic acid binding / zinc ion binding	No reported	TRIAE_CS42_U_TGACv1_669171_AA2155140.2
S7B_5745497	P	Nucleic acid binding / RNA - DNA Hybrid ribonuclease activity	DNA integration	TRIAE_CS42_7BL_TGACv1_576794_AA1855180

Table 4. Very significant ($p < 0.0005$) and significant ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14) and St. Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

Very significant and consistent p < 0.0005

SNP	Trait	Environment	P. value	MAF	R^2 [a]	Allelic Effect [b]
S6B_309688673	Ca	SJ13	3.2E-05	0.07	0.23	-0.03
		SJ14	1.7E-04	0.06	0.11	-0.03
S6B_309688674	Ca	SJ13	3.2E-05	0.07	0.23	-0.03
		SJ14	1.7E-04	0.06	0.11	-0.03
S7D_579146731	Na	SJ13	5.8E-05	0.06	0.13	38.33
		SJ14	2.3E-04	0.06	0.15	22.84
S1A_379438389	Fe	SJ14	4.5E-04	0.40	0.08	11.34
		ST13	2.6E-04	0.40	0.10	-0.01
S1A_372238511	Fe	SJ14	3.3E-04	0.41	0.08	11.65
		ST13	2.2E-04	0.41	0.10	-0.01
S1A_376790586	Fe	SJ14	2.5E-04	0.40	0.08	11.86
		ST13	4.7E-04	0.40	0.10	-0.01
S1A_373360656	Fe	SJ14	2.1E-04	0.40	0.08	12.01
		ST13	3.3E-04	0.40	0.10	-0.01
S1A_373812006	Fe	SJ14	2.0E-04	0.41	0.08	12.17
		ST13	2.9E-04	0.40	0.10	-0.01
S1A_375844331	Fe	SJ14	1.5E-04	0.41	0.09	12.27
		ST13	4.7E-04	0.40	0.10	-0.01
S1A_382337060	Fe	SJ14	1.1E-04	0.41	0.09	12.57
		ST13	3.6E-04	0.41	0.10	-0.01
S1A_373846553	Fe	SJ14	1.1E-04	0.40	0.09	12.55
		ST13	5.0E-04	0.40	0.10	-0.01
S1A_380869075	Fe	SJ14	7.3E-05	0.41	0.09	12.95
		ST13	4.0E-04	0.41	0.10	-0.01
S1B_599660615	Na	ST13	1.7E-05	0.18	0.17	165.84
		SJ14	1.0E-04	0.18	0.16	14.76
S5B_588714935	Na	ST13	1.5E-04	0.09	0.15	197.80
		SJ14	3.9E-04	0.09	0.14	18.06
S5B_588714936	Na	ST13	1.5E-04	0.09	0.15	197.80
		SJ14	3.9E-04	0.09	0.14	18.06
S7D_528997111	Zn	ST13	2.2E-05	0.11	0.13	1.97
		SJ14	1.2E-05	0.11	0.21	1.08
S1B_683422400	Zn	ST13	1.9E-05	0.11	0.13	1.94
		SJ14	4.6E-05	0.11	0.20	0.97
S7B_567704027	Zn	ST13	1.8E-05	0.11	0.13	1.99
		SJ14	4.7E-04	0.11	0.18	0.86
S7B_568012679	Zn	ST13	1.8E-05	0.11	0.13	1.99
		SJ14	4.7E-04	0.11	0.18	0.86
SUN_334466963	Al	SJ14	3.3E-05	0.05	0.10	13.60
		ST14	3.4E-05	0.05	0.10	35.90
SUN_334466979	Al	SJ14	3.3E-05	0.05	0.10	13.60
		ST14	3.4E-05	0.05	0.10	35.90
SUN_221478096	Al	SJ14	2.3E-04	0.05	0.08	14.90
		ST14	1.9E-04	0.05	0.17	214.72
S5A_3619242	Fe	ST14	3.0E-04	0.10	0.08	21.16
		ST13	3.6E-04	0.10	0.11	1.64

Table 4 (Cont.). Very significant ($p < 0.0005$) and significant ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14) and St. Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Environment	P. value	MAF	R^{2a}	Allelic Effect [b]
Significant and consistent p < 0.001						
S1A_302553240	Na	SJ14	9.2E-04	0.08	0.14	18.43
		ST13	8.1E-04	0.08	0.14	187.63
S3A_720852820	Na	SJ14	5.3E-04	0.06	0.14	18.87
		ST13	6.2E-04	0.06	0.14	189.26
S3B_50499819	Mg	SJ13	9.5E-04	0.06	0.13	0.01
		SJ14	5.6E-04	0.06	0.09	0.01
S6D_466814084	K	SJ14	7.7E-04	0.23	0.16	0.10
		ST13	5.6E-04	0.23	0.11	-0.11
S6D_466814117	K	SJ14	7.7E-04	0.23	0.16	0.10
		ST13	5.6E-04	0.23	0.11	-0.11
S7B_564915089	Zn	SJ14	8.1E-04	0.19	0.18	0.72
		ST13	5.6E-04	0.19	0.10	1.35
S7B_564915141	Zn	SJ14	8.1E-04	0.19	0.18	0.72
		ST13	5.6E-04	0.19	0.10	1.35
S7B_5745497	P	ST13	8.8E-04	0.46	0.09	0.01
		ST14	6.9E-04	0.46	0.11	0.01

[a] Reflect the phenotypic variation explained by the marker, R^2 of the model with SNP calculated in GAPIT package in R

[b] Allelic effects with respect to the minor allele

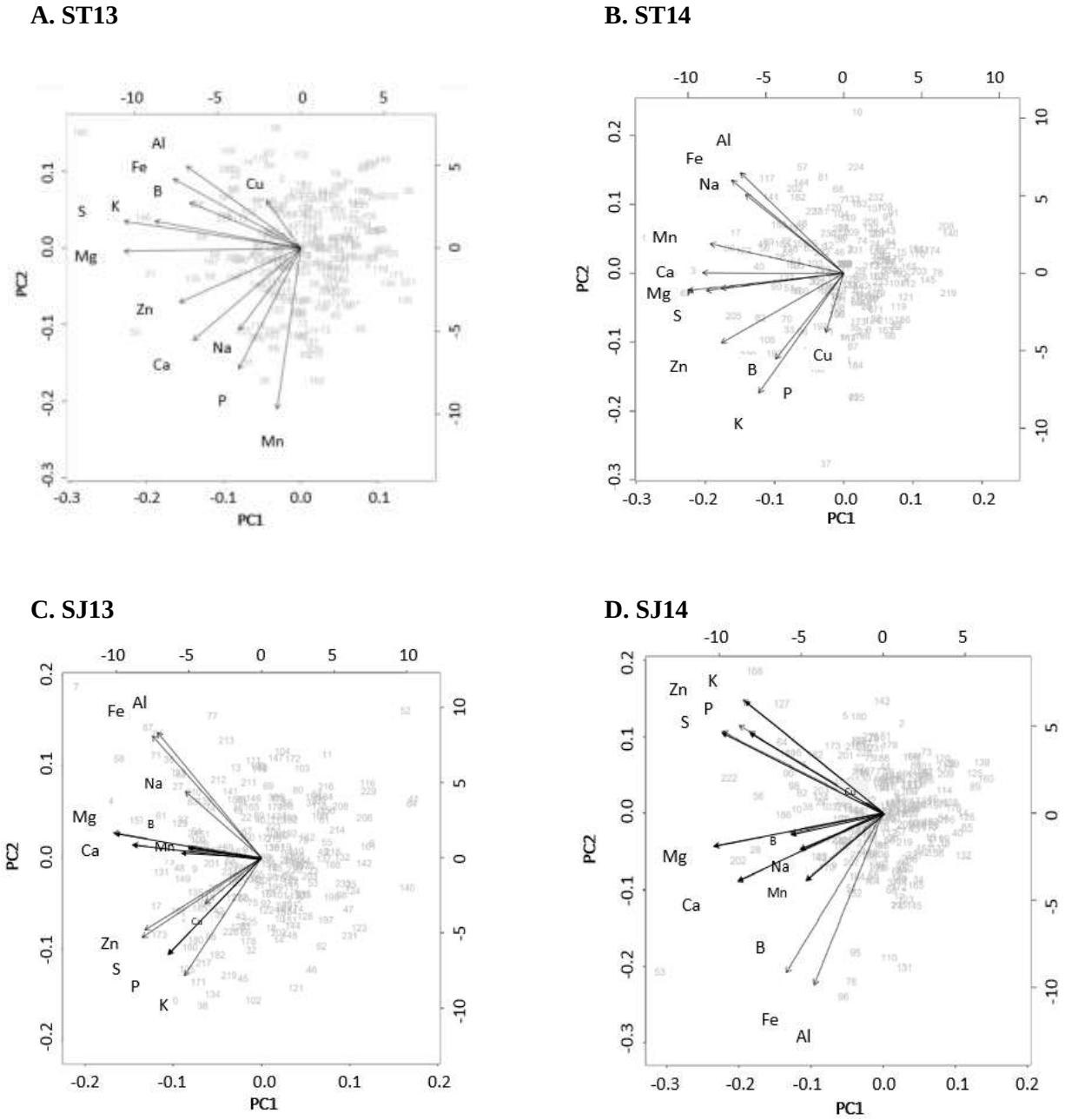

Figure 1. PCA bi-plots of the genetic correlation of micro and macronutrient concentrations across different site-years for the soft winter wheat association panel (AP). **A**. **ST13:** Stuttgart, AR 2013; B**. ST14:** Stuttgart, AR 2014; **C**. **SJ13:** St. Joseph, LA 2013; D**. SJ14:** St. Joseph, LA 2014.

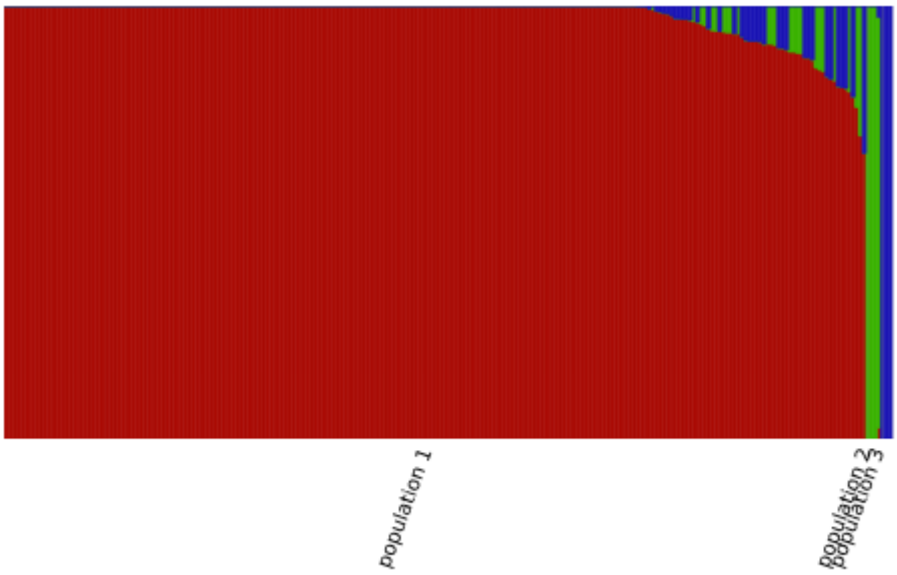

Figure 2. Population structure based on 240 genotypes and 92,916 SNP markers. Each colored region represents a subpopulation.

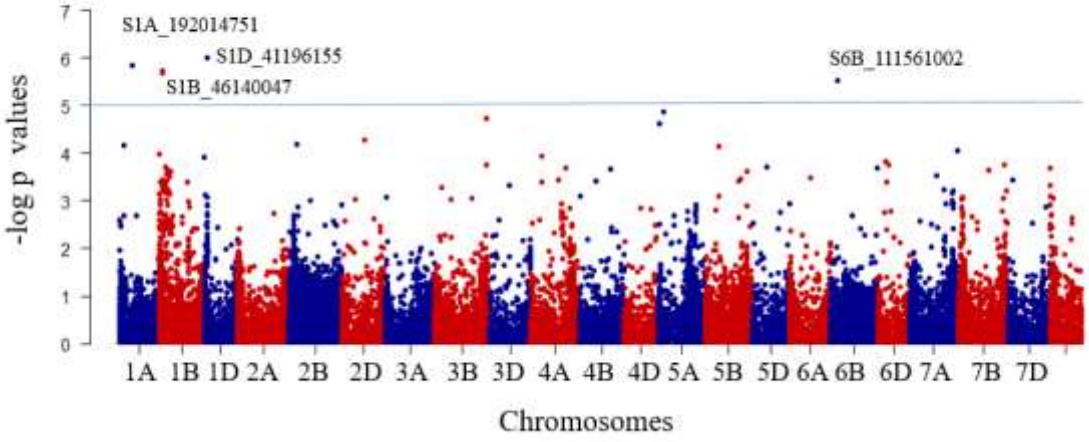

Figure 3. Manhattan plot showing genome-wide SNP loci associated with [Fe] in ST13 under a *K-Q* model. Horizontal line represents the significant threshold by which markers were considered associated with a trait ($p < 0.00001$; ~5.0).

References

Acuña-Galindo MA, Mason RE, Subramanian NK, Hays DB (2015) Meta-analysis of wheat QTL regions associated with adaptation to drought and heat stress. Crop Science 55:477-492

Ahmed F, Rafii MY, Ismail MR, Juraimi AS, Rahim HA, Asfaliza R, Latif MA (2013) Waterlogging tolerance of crops: breeding, mechanism of tolerance, molecular approaches, and future prospects. BioMed research international 2013:963525

Arguello MN, Mason RE, Roberts TL, Subramanian N, Acuña A, Addison CK, Lozada DN, Miller RG, Gbur E (2016) Performance of soft red winter wheat subjected to field soil waterlogging: Grain yield and yield components. Field Crops Research 194:57-64

Bailey-Serres J, Lee SC, Brinton E (2012) Waterproofing crops: effective flooding survival strategies. Plant Physiology 160:1698-1709

Ballesteros DC, Mason RE, Addison CK, Acuna MA, Arguello MN, Subramanian N, Miller RG, Sater H, Gbur EE, Miller D, Griffey CA, Barnett MD, Tucker D (2015) Tolerance of wheat to vegetative stage soil waterlogging is conditioned by both constitutive and adaptive QTL. Euphytica 201:329-343

Benson J, Brown-Guedira G, Paul Murphy J, Sneller C (2012) Population structure, linkage disequilibrium, and genetic diversity in soft winter wheat enriched for fusarium head blight resistance. The Plant Genome 5:71-80

Bentley A, Scutari M, Gosman N, Faure S, Bedford F, Howell P, Cockram J, Rose G, Barber T, Irigoyen J, Horsnell R, Pumfrey C, Winnie E, Schacht J, Beauchêne K, Praud S, Greenland A, Balding D, Mackay I (2014) Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. Theoretical and Applied Genetics 127:2619-2633

Bordes J, Goudemand E, Duchalais L, Chevarin L, Oury FX, Heumez E, Lapierre A, Perretant MR, Rolland B, Beghin D, Laurent V, Le Gouis J, Storlie E, Robert O, Charmet G (2014) Genome-wide association mapping of three important traits using bread wheat elite breeding populations. Mol Breeding 33:755-768

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635

Collaku A, Harrison SA (2002) Losses in wheat due to waterlogging. Crop Science 42:444-450

Collaku A, Harrison SA (2005) Heritability of waterlogging tolerance in wheat. Crop Science 45:722-727

Dai J, Bai G, Zhang D, Hong D (2013) Validation of quantitative trait loci for aluminum tolerance in Chinese wheat landrace FSW. Euphytica 192:171-179

Delhaize E, Craig S, Beaton CD, Bennet RJ, Jagadish VC, Randall PJ (1993) Aluminum tolerance in wheat (*Triticum aestivum* L.) (I. Uptake and distribution of aluminum in root apices). Plant Physiology 103:685-693

Donohue SJ, Aho. DW (1992) Determination of P, K, Ca, Mg, Mn, Fe, Al, B, Cu, and Zn in plant tissue by inductively coupled plasma (ICP) emission spectroscopy. C.O. Plank (ed.). Plant Analysis Reference Procedures for the Southern Region of the United States. Georgia Agric. Expt. Sta., Southern Coop. Ser. Bul. 368.

Drew M, Lynch J (1980) Soil anaerobiosis, microorganisms, and root function. Annual Review of Phytopathology 18:37 - 66

Flint-Garcia SA, Thornsberry JM, S E, IV B (2003) Structure of linkage disequilibrium in plants Annual review of plant biology 54:357-374

Foy CD, Fleming AL (1982) Aluminum tolerances of two wheat genotypes related to nitrate reductase activities. Journal of Plant Nutrition 5:1313-1333

Francis RM (2017) POPHELPER: An R package and web app to analyse and visualize population structure. Mol Ecol Resour 17:27-32

Froese PS, Carter AH (2016) Single nucleotide polymorphisms in the wheat genome associated with tolerance of acidic soils and aluminum toxicity. Crop Science 56:1662-1677

Garbisu C, Alkorta I (2001) Phytoextraction: a cost-effective plant-based technology for the removal of metals from the environment. Bioresource technology 77:229-236

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461-485

Jefferies SP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam AKMR, Langridge P, Chalmers KJ (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theoretical and Applied Genetics 101:767-777

Hamblin MT, Buckler ES, Jannink JL (2011) Population genetics of genomics-based crop improvement methods. Trends Genet 27:98-106

Harborne JB (1989) Introduction to ecological biochemistry. Academic Press Harcourt Brace Jovanovich Publishers, New York

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annual review of plant biology 65:531-551

Jefferies SP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam AKMR, Langridge P, Chalmers KJ (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theoretical and Applied Genetics 101:767-777

Kaler AS, Ray JD, Schapaugh WT, King CA, Purcell LC (2017) Genome-wide association mapping of canopy wilting in diverse soybean genotypes. Theoretical and Applied Genetics 130:2203-2217

Khabaz-Saberi H, Rengel Z (2010) Aluminum, manganese, and iron tolerance improves performance of wheat genotypes in waterlogged acidic soils. Journal of Plant Nutrition and Soil Science 173:461-468

Khabaz-Saberi H, Setter TL, Waters I (2005) Waterlogging induces high to toxic concentrations of iron, aluminum, and manganese in wheat varieties on acidic soil. Journal of Plant Nutrition 29:899-911

Kirk GJD, Solivas JL, Alberto MC (2003) Effects of flooding and redox conditions on solute diffusion in soil. European Journal of Soil Science 54:617-624

Kochian LV, Hoekenga OA, Piñeros MA (2004) How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. Annu Rev Plant Biol 55:459-493

Kochian LV, Piñeros MA, Liu J, Magalhaes JV (2015) Plant adaptation to acid soils: The molecular basis for crop aluminum resistance. Annual review of plant biology 66:571-598

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397-2399

Ma HX, Bai GH, Carver BF, Zhou LL (2005) Molecular mapping of a quantitative trait locus for aluminum tolerance in wheat cultivar Atlas 66. Theor Appl Genet 112:51-57

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends in Plant Science 12:57-63

Manickavelu A, Hattori T, Yamaoka S, Yoshimura K, Kondou Y, Onogi A, Matsui M, Iwata H, Ban T (2017) Genetic nature of elemental contents in wheat grains and its genomic prediction: toward the effective use of wheat landraces from Afghanistan. PLoS One 12 (1): e0169416

Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S (2015) LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3 (Bethesda, Md) 5:2383-2390

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. The Plant Cell Online 21:2194-2202

Navakode S, Neumann K, Kobiljski B, Lohwasser U, Börner A (2014) Genome wide association mapping to identify aluminium tolerance loci in bread wheat. Euphytica 198:401-411

Oyiga BC, Sharma RC, Baum M, Ogbonnaya FC, Léon J, Ballvora A (2017) Allelic variations and differential expressions detected at quantitative trait loci for salt stress tolerance in wheat. Plant, cell & environment; 41(5):919-935. doi: 10.1111/pce.12898

Ponnamperuma FN (1972) The chemistry of subemerged soils. Advances in agronomy 24 29-96.

Raj A, Stephens M, Pritchard JK (2014) fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genetics 197:573-589

Raman H, Stodart B, Ryan PR, Delhaize E, Emebiri L, Raman R, Coombes N, Milgate A (2010) Genome-wide association analyses of common wheat (*Triticum aestivum* L.) germplasm identifies multiple loci for aluminium resistance. Genome 53:957-966

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the Bo1 boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451-461

Setter T, Waters I, Sharma S, Singh KN, Kulshreshtha N, Yaduvanshi NPS, Ram PC, Singh BN, Rane J, McDonald G, Khabaz-Saberi H, Biddulph T, Wilson R, Barclay I, McLean R, Cakir M (2008) Review of wheat improvement for waterlogging tolerance in Australia and India: The importance of anaerobiosis and element toxicities associated with different soils 103:221-235

Setter TL, Waters I (2003) Review of prospects for germplasm improvement for waterlogging tolerance in wheat, barley and oats. Plant and Soil 253:1-34

Setter TL, Waters I, Sharma SK, Singh KN, Kulshreshtha N, Yaduvanshi NP, Ram PC, Singh BN, Rane J, McDonald G, Khabaz-Saberi H, Biddulph TB, Wilson R, Barclay I, McLean R, Cakir M (2009) Review of wheat improvement for waterlogging tolerance in Australia and India: the importance of anaerobiosis and element toxicities associated with different soils. Ann Bot 103:221-235

St. Burgos M, Messmer MM, Stamp P, Schmid JE (2001) Flooding tolerance of spelt (*Triticum spelta* L.) compared to wheat (*Triticum aestivum* L.). A physiological and genetic approach. Euphytica 122:287-295

Wang Y, Stass A, Horst WJ (2004) Apoplastic binding of aluminum is involved in silicon-induced amelioration of aluminum toxicity in maize. Plant Physiol 136:3762-3770

Yang JL, Li YY, Zhang YJ, Zhang SS, Wu YR, Wu P, Zheng SJ (2008) Cell wall polysaccharides are specifically involved in the exclusion of aluminum from the rice root apex. Plant Physiology 146:602-611

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics 38:203-208

Yu M, Chen G-Y (2013) Conditional QTL mapping for waterlogging tolerance in two RILs populations of wheat. SpringerPlus 2:245

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Gen 1:5-20

CHAPTER IV

EVALUATION OF THE WHEAT TRANSCRIPTOME UNDER WATERLOGGING CONDITIONS USING RNA-SEQ ANALYSIS

Abstract

Waterlogging is a major constraint to global wheat (*Triticum aestivum* L.) production as it disrupts physiological processes such as photosynthesis, respiration and carbohydrate metabolism, resulting in an energy crisis that impacts yield production. The identification of genes, gene networks and biological crosstalk of different stress pathways is required for a holistic approach to crop improvement. RNA-seq analysis was used to evaluate the gene expression of two wheat genotypes under waterlogged and non-waterlogged conditions. Two "iconic" US soft red winter wheat cultivars, 'Pioneer Brand 26R61' and 'AGS 2000' were subjected to waterlogging stress 24 days after sowing at the tillering stage, in a growth chamber and plant leaf tissue samples were collected after 24 hours of stress imposition. RNA-seq analysis was carried out using 2x100 bp paired-ends sequencing on the HiSeq2500 platform. Around 300 million pair-end sequence reads were generated, covering approximately 16 Gb of the wheat transcriptome. In total around 60,414 CDS were obtained for AGS200 and 26R61, respectively and 58,753 expressed genes were observed across both cultivars and treatments. Gene annotation was successful for 117 significantly expressed genes. These genes were associated with oxidation-reduction reactions, photosynthesis/light reaction, metabolism, thiamine biosynthetic process and carbohydrate metabolic. Transcript levels of the genes *Traes_5BL_56C52F2D9.1, Traes_2AL_6F3A92BF2.1* and *Traes_5DS_7FD4F7284.1* were confirmed using qRT-PCR.

Introduction

Waterlogging (WL) is a significant constraint to wheat production (Ding et al. 2018) and changes in precipitation patterns due to global climate change have resulted in a higher frequency of this stress during crop growing seasons (Bailey-Serres et al. 2012; Tamang and Fukao 2015). The development of new tolerant wheat varieties is vital to meet future production demands (Herzog et al. 2016). Soil waterlogging results in anoxia and hypoxia due to decreased availability of soil oxygen (Ponnamperuma 1972). This reduction in oxygen impairs gas diffusion and increases production of toxic micro and macro elements available for plant uptake (Setter et al. 2009). This results in changes at the molecular and metabolic levels in the plant to suppress damage resulting from an energy crises due to a reduction in the production of adenosine triphosphate (ATP) (Gupta et al. 2009).

Abiotic stresses trigger plant adaptive strategies through modification of cellular and molecular processes that lead to changes in growth and development and (Ahuja. et al. 2010). These modifications are accompanied by modifications in gene expression and protein and metabolite production (Bokhari et al. 2007; Chae et al. 2009; Urano et al. 2009). Klok et al. (2002) observed complex responses in gene expression due to low oxygen levels in Arabidopsis (*Arabidopsis thaliana*), with genes involved in nitrogen metabolism, including nitrate reductase, glutamate dehydrogenase and glutamate decarboxylase being overexpressed. Similarly, in the analysis of Arabidopsis and rice (*Oryza sativa* L.) transcriptomes a change in the expression of 5 to 10% of the total genes evaluated was observed as a result of low oxygen conditions (Klok et al. 2002; Lasanthi-Kudahettige et al. 2007). Ethylene accumulation in plant tissues under WL conditions has been associated with WL tolerance in rice by inducing gene expression, including the *Submergence 1* locus *(SUB1)* (Xu et al. 2006) which reduces plant growth rate and the

snorkel (SK) (Hattori et al. 2009) locus which increases stem elongation rate, thus avoiding total plant submergence. Overall, genes expression is dependent on the type, duration and severity of the stress (Mahajan and Tuteja 2005).

RNA sequencing (RNA-seq) allows for both discovery and gene quantification in a single experiment (Conesa et al. 2016) (Adams et al. 1991). RNA-Seq analysis can be applied in both organisms with well-annotated genomes through direct mapping of sequences, or using a *de novo* approach in which contigs are assembled and then mapped to the transcriptome (Conesa et al. 2016). Chen et al. (2016) evaluated differentially expressed genes (DEG) of soybean plants under drought and waterlogging conditions and found down regulation of genes involved in photosynthetic and chlorophyll synthesis under both stresses. Du et al. (2017) used bulked segregant RNA and found 431 DEGs under waterlogging stress in maize (*Zea mays* L). In addition, *GRMZM2G055704* was identified as a candidate gene on chromosome 1 where a previous QTL associated with waterlogging tolerance was reported. Bhardwaj et al. (2015) evaluated gene regulation under heat and drought conditions in *Brassica juncea* seedlings, identifying more than 97,000 transcripts of which 19,000 were differentially regulated and associated with different metabolic pathways such as purine metabolism, amino sugar and nucleotide sugar metabolism and lipopolysaccharide biosynthesis.

Despite of the impact of waterlogging stress on wheat production, there are few studies evaluating the molecular basis and gene networks of the waterlogging stress response in wheat. The objective of this study was to evaluate the wheat transcriptome response under waterlogged and non-waterlogged conditions in two “iconic” varieties of soft red winter wheat adapted to the southeastern US, an area prone to annual soil waterlogging. Our results provide new insights into

wheat's response to waterlogging stress and will aid in future wheat improvement through a better understanding of the gene networks involved.

Materials and methods

Plant material and experimental design

Wheat cultivars 'Pioneer Brand 26R61' and 'AGS 2000' were evaluated under growth chamber soil waterlogging and non-waterlogging conditions. Evaluation of these cultivars under field waterlogging stress, as reported in Chapter 2 of this dissertation, showed AGS200 to have lower elemental accumulation and higher yield components compared to Pioneer 26R61. For the WL stress, cultivars were sown in a sandy loam soil (56.9% sand, 37% silt, and 6.1% clay) at five seeds per 1.65 liter tree pot (10 cm wide x 24 cm tall) and the pots were placed in 50 liter plastic tubs. The growth chamber was maintained at an average temperature of 20°C/18°C day/night cycle with 16 hour days and a light intensity of 600 mmol m^{-2} s^{-1} PAR. A waterlogging (WL) treatment was applied 22 days after sowing to waterlogging Pioneer 26R61 (WP) and waterlogging AGS200 (WA) cultivars. Control plants from Pioneer 26R61 (CP) and AGS200 (CA) were maintained in plastic tubs with no waterlogging (non-WL) treatment to maintain normal growth conditions. Plant leaf tissue was collected 24 hours after stress imposition as it has been shown that a reduction of root oxygen level and thus gene expression by WL conditions begins after four hours of stress imposition (Christianson et al. 2009). Moreover, ATP levels at cellular level can be reduced by 50% after two hours of low oxygen (Branco-Price et al. 2008). Immediately after collection, plant tissues were placed in liquid nitrogen and stored in -80 ºC.

RNA isolation and cDNA synthesis

Total leaf RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, CA). After sample quantification, 1 ug of total RNA of the four samples with one replication for each treatment was used to synthesize cDNA using the Reverse transcription System (Promega, WI). Sequences of housekeeping primers were obtained from Christianson et al. (2009) for *ADH2BF* (*gac ctc tac ctt cag cga gta*) and *ADH2BR* (*gag ata cca cag ctg aga acac*) and Tenea et al. (2011) for primers *Hous1* (*cac cgg ccc agt gat ctt*) and *Hous1R* (*aag ggc gtc tgc tcc aact*). Quantitative real-time PCR was performed using the GoTaq qPCR Master Mix (Promega, WI). using program conditions 95 °C for 2min; 39 cycles of 95 °C for 15 sec, 59 °C for 1min; 65 °C for 1min and 95 °C for 5 sec. RNA integrity and quantification was performed using the Bio-rad Experion System (Bio-rad Laboratories, Inc. Hercules, CA).

Sample sequencing and assembly

After RNA quantification, a validation was conduction on the Agilent Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany). Eight total samples consisting of two replications of non-waterlogging conditions (control) and two waterlogging condition (stress) from each cultivar were sent to the Research Technology Support Facility at Michigan State University where libraries were prepared using the Illumina TruSeq Stranded mRNA Library Kit (LT). After validation and quantitation the libraries were pooled and loaded on one lane of an Illumina HiSeq 2500 Rapid Run flow cell (v1) (Illumina® San Diego, CA). Sequencing was carried out in a 2x100bp (PE100) format using Illumina Rapid SBS reagents. Base calling was done by Illumina Real Time Analysis (RTA) v1.17.21.3 and output of RTA de-multiplexed and converted to FastQ format with Illumina Bcl2fastq v1.8.4. Around 298,360,064 million pair-ends read were obtained from RNA sequencing under waterlogging and non- waterlogging conditions

combining both long reads (200 bp) and high coverage (3.8 million reads, reaching a total of 1,369 megabases). Data was read using Bowtie2 (Langmead and Salzberg 2012). Coding sequences (CDS) from the *Tritium aestivum* genome consortium IWGSP version 1.23.cdna were used to create an index file to map the RNA-seq reads. The reads were parsed and counted based on CDS tags divided by length of the CDS to reads per kilobase of transcript per million mapped reads (RPKM). Alignment data was then filtered and CDS with only 10 or more reads were used for analysis. In addition, annotated genomes (A, B and D) of *T. aestivum* coding sequences (CDS) from the database *EnsemblPlants* were used as a reference genome for data alignment using SAMtools (Li et al. 2009).

Gene Validation

Validation of the selected reads was performed using quantitative real-time PCR (qRT-PCR). Six significant annotated CDS were chosen for gene validation and primer design was performed using the PrimerQuest Tool (www.idtdna.com/PrimerQuest/Home/Index). Samples for qRT-PCR were prepare following the protocol from GoTaq qPCR Master Mix (Promega) where 1 µL cDNA (100ng/ul), 0.5 µL (10 µM) reverse primer, 0.5 µL (10 µM) forward , 5ul of GoTaq Maxter Mix and 3uL of ddH_2O for a 10uL reaction were used. Amplification was done using the Bio-Rad CFX96 Real Time PCR System (Bio-Rad Laboratories, Inc., Hercules, CA, USA). Reaction conditions for thermal cycling were: 95 °C for 2min; 39 cycles of 95 °C for 15 sec, 59 °C for 1min; 65 °C for 1min and 95 °C for 5 sec, followed by a melt-curve analysis to confirm single PCR product amplification. Analysis of the data was performed using the CFX Manager V.1.5 software. Gene expression calculation was done as described by Livak and Schmittgen (2001) where cycle threshold (CT) values are calculated to detect and quantify fluorescent signal. CT values were calculated as follows,

ΔCT control (non-WL) = [CT value control– (CT internal control (housekeeping genes)), and

ΔCT stress (WL) = [CT value stress– (CT internal control (housekeeping genes))

ΔΔCT = [ΔCT stress (WL) − ΔCT control (non-WL].

Relative gene expression (Y) = Y (Individual sample) = 2 $^{(-\Delta\Delta CT)}$

Results

Total number of transcripts and CDS identified

Using the 2x100 bp paired end sequencing, approximately 300 million pair end reads were generated, which covered approximately 16 Gb of the wheat transcriptome. After alignment with the Chinese spring wheat (*Triticum aestivum* L.) genome available in the *EnsemblPlants*, a total 100,717 CDS were identified. After data filtering for CDS with ten or more reads 64,911 and 60,414 CDS were obtained for AGS200 and 26R61, respectively. A similar number of expressed genes were observed across treatments and cultivars, CA (63,616), CP (62,745), WA (64,745) and WP (65280). (Fig. 1). The number and distribution of CDS were consistent with those reported in the *Triticum aestivum* L. database *EnsemblPlants*, with the largest number of CDS reported for the Group 2 chromosomes and the lowest on the Group 1 and Group 6.

Gene expression

A total of 58,753 expressed genes were observed across both cultivars and treatments (Fig. 2). Waterlogging stress when compared with non WL resulted in an increase of 1129 and 2535 differentially expressed genes (DEG) in AGS200 and 26R61, respectively. Under WL conditions, the greatest number of expressed genes was observed for 26R61 with 1824 expressed genes compared with 1085 for AGS2000. Based on a level of expression of $\leq$ -2.0 or $\geq$ 2.0, 1020 genes were determined to be significantly expressed under waterlogging stress in 26R61 with

447 and 576 genes repressed and induced, respectively. Likewise, 554 genes were significantly expressed in AGS200, with 301 and 254 genes being repressed and induced, respectively.

Gene Annotation

Gene ontology identified 3142 and 3086 gene annotations for AGS200 and 26R61, respectively, or approximately 5% of the total CDS assembled in this study have a reported molecular or biological function. Cytosine-specific methyl transferase was regarded as the molecular function with most number of CDS at 247 (Fig. 3). On the other hand, the top five biological upregulated processes with most number of CDS were oxidation-reduction reaction, photosynthesis/light reaction, metabolism, thiamine biosynthetic process and carbohydrate metabolic (Fig. 4). Only 117 of the significantly expressed genes had functional annotation, including twenty-six for AGS200 (Table 1) and 91 for 26R61 (Table 2). Annotation hits included ribosomal proteins, cytochrome b6, HistoneH2A, serine/threonine protein kinases and eukaryotic factor translation initiation factor (Table 2).

Gene validation

To validate the results of RNA-seq, six annotated and significantly expressed genes were chosen for quantitative real-time PCR (qRT-PCR). These included *Traes_2AL_6F3A92BF2.1, Traes_5DS_7FD4F7284.1, Traes_4AL_CC989B4A5.1* for AGS200 and *Traes_5BL_D526A626E.2, Traes_5BL_56C52F2D9.1* and *Traes_2AL_520618712.1* for 26R61. (Table 3). Relative expression was in agreement for three of the six genes, with the other three showing non-amplification. *Traes_5BL_56C52F2D9.1*, which has an annotated function as ribulose bisphosphate carboxylase small chain reactions showed higher expression under non-WL condition. Similarly, an increase of one fold was observed for *Traes_2AL_6F3A92BF2.1* annotated as a non-specific lipid-transfer protein in cultivar 26R61. Under WL conditions an

increase of almost five fold was observed for *Traes_5DS_7FD4F7284.1*, which is a potential phosphoglycerate kinase (PGK).

Discussion

CDS identified and RNA-seq validation

We evaluated two winter wheat cultivars under WL and non-WL conditions using RNA-seq to identify differentially expressed genes. A total of 100,117 CDS were assembled and were in agreement in terms of distribution with those reported in the database *EnsemblPlants.* Results from RNA-transcriptome assembly were in agreement with RNA-seq results confirming the correct assembly of the transcriptome.

Gene Expression and annotation

Gene annotation has been improved for model crops, however annotation accuracy continues to be a difficult task, even for genes sequences present across species for which a function has not yet been characterized (Bolger et al. 2018). For example, although it is one of the most well annotated genomes, around 20% of Arabidopsis genes have not been experimentally characterized (TAIR, http://arabidopsis.org). Hexaploid wheat, consisting of three genomes and around 80% of the genome being repetitive sequences (Moore et al. 1995) poses many challenges for successful gene annotation. Despite the recent release of a complete draft wheat genome, a high quality gene annotation has not yet been produced (Clavijo et al. 2017).

In this study, gene annotations were available for 6,228 of the 58,000 expressed genes in this study. This included 3142 and 3086 for AGS200 and 26R61 respectively. Under WL conditions cultivar 26R61 presented a higher number of expressed genes compared to AGS2000. Several sequences were found to be present on several chromosomes but not in the specific chromosome were the CDS was sequenced, pointing out the repetitive gene sequences in the

wheat genome. Nevertheless, we were able to identify important CDS associated with key molecular and biological processes involved in the response to abiotic stress. For example, cytosine-specific methyl transferase was the molecular function commonly identified across 247 CDS. DNA methyltransferases regulate epigenetic expression in plants, catalyzing the methylation of the cytosine DNA residues. In addition, they regulates the chromatin structure in plants which determines plant adaptation under environmental factors (Wada 2005) and gene expression during plant development (Finnegan and Kovac 2000). In rice, suppression of genes associated with cytosine DNA methyltransferases was observed in response to cold and salt stress (Rita et al. 2009). Similar results were obtained by Steward et al. (2002) in which an interaction was observed between chromatin elements and DNA methylation. Additionally, modification of DNA methylation levels were observed in cotton (*Gossypium hirsutum*) under cold stress conditions (Fan et al. 2013).

Serine/Threonine proteins are phosphorylation proteins involved in the regulation of adaptive stress, for which studies showing their functionality in plant adaptation under cold (Vazquez-Tello et al. 1998), drought (Yu et al. 2003) and biotic (MacKintosh et al. 1994) stresses have been reported, although not yet shown for waterlogging stress.

PGK which is part of the glycolytic and fermentation pathway was shown to be upregulated under WL (Alam et al. 2010; Kosová et al. 2011). Similarly Qi et al. (2012) observed an increase in transcript number of this enzyme under waterlogging conditions in cucumber (*Cucumis sativus* L.). However, contrasting results were observed by Gu et al. (2011) in which evaluation of *Chrysanthemum* L. found stable expression of this gene under control and stress (heat and waterlogging) conditions. On the other hand, down regulation of ribulose bisphosphate carboxylase (RUBISCO) which fixes carbon dioxide during photosynthesis and

uses oxygen during photorespiration was observed and is in agreement with Sage et al. (1988) who observed a decrease of RUBISCO activity in response to accumulation of carbon dioxide which has been shown to increase under flooding conditions (Yasuo et al.1956). Finally, non-specific lipid-proteins are present in a high amount in plants have been associated with resistance to abiotic stress (Liu et al. 2015) and play an important role in membrane stabilization (Hincha et al. 1997).

Around 306 CDS were associated with ribosomal proteins, however, gene expression was not consistent, showing significant fold-change in both upregulated (7.37) and downregulated (-5.76) genes. Ribosomal proteins are involved in protein synthesis, playing a key role in plant metabolic processes, cell division, plant growth and stress response/tolerance (Whittle and Krochko 2009, Saha et al. 2017). In addition, as being part of the cytoplasmic ribosomal structure, around 80 ribosomal proteins and 249 genes have been associated with their expression in Arabidopsis (Barakat et al. 2001). Evaluating maize under abiotic stress (UV-B) Falcone Ferreyra et al. (2010) observed a suppression of these proteins, affecting several processes such as growth and development and a reduction in productivity. In soybean (*Glycine max*) down regulation of these proteins was observed under flooding (Oh et al. 2014b) but an increase when soybean roots were exposed to exogenous calcium under flooded conditions, affecting protein synthesis and degradation, cell wall integrity, DNA synthesis and hormone metabolism (Oh et al. 2014a).

Conclusion

This study is a global evaluation of gene expression under waterlogging stress and is resource for identification of tolerant genes under waterlogging stress. One hundred and seventeen CDS were

associated with key biological processes associated with responses to abiotic stresses and serve as targets for the development of waterlogging tolerant wheat lines.

Table 1. Significant differentially expressed genes evaluated in cultivar AGS2000 under waterlogging vs. non-waterlogging conditions.

Upregulated		
CDS	**Annotation**	**Fold change**
Traes_2DL_3B2B9300C.1	GrpE protein homolog	2.39
Traes_5DS_4F07516F7.1	30S ribosomal protein S15, chloroplastic	2.37
Traes_2AL_DB64E18A1.1	Ribulose bisphosphate carboxylase small chain	2.35
Traes_1AL_93E51B797.1	Ribosomal protein S7	2.30
Traes_2DL_6E8EAFA1D.1	4-hydroxy-tetrahydrodipicolinate synthase 2, chloroplastic	2.16
Traes_1DS_F3F7B8AFE1.6	30S ribosomal protein S3, chloroplastic	2.11
Traes_2BL_C433BB333.3	Cytochrome b6-f complex subunit 6	2.08
Traes_2DL_1F24E4FF8.1	Reticulon-like protein	2.05
Traes_7DS_CCFC4F086.1	Ribonucleoside-diphosphate reductase	2.03
Downregulated		
CDS	**Annotation**	**Fold change**
Traes_2AL_6F3A92BF2.1	Non-specific lipid-transfer protein	-3.53
Traes_5DS_3186C8F7D.1	50S ribosomal protein L20, chloroplastic	-3.40
Traes_5DS_7FD4F7284.1	Phosphoglycerate kinase	-3.21
Traes_6DL_878B1A04B.1	Phosphoglycerate kinase	-2.51
Traes_3AL_655060842.1	50S ribosomal protein L20	-2.45
Traes_5AS_7D519210E.2	Eukaryotic translation initiation factor 3 subunit C	-2.39
Traes_6DL_2EC6B5AB9.1	30S ribosomal protein S18, chloroplastic	-2.34
Traes_4DS_DBAA2CC451.2	30S ribosomal protein S11, chloroplastic\x3b DNA-directed RNA polymerase subunit alpha	-2.28
Traes_2AL_63B2E5BC2.1	30S ribosomal protein S15	-2.14
Traes_2BL_B877F41B8.1	Actin-related protein 2/3 complex subunit 5	-2.11
Traes_1DS_1360CF476.1	40S ribosomal protein S8	-2.10
Traes_4AL_0D6C3F8F2.2	30S ribosomal protein S18, chloroplastic	-2.06
Traes_3B_21DEF07D2.2	Pyruvate kinase	-2.05
Traes_4AL_A0BA8C598.1	DNA-directed RNA polymerase	-2.03
Traes_7AS_BF53A95BA.1	Auxin response factor	-2.01
Traes_1AS_C9FF7774B.2	TATA-box-binding protein 1	-1.99
EPlTAET00000009103	Metazoan signal recognition particle RNA	-1.96

Table 2. Significant differentially expressed genes evaluated in cultivar pioneer 26R61 under waterlogging vs. non-waterlogging conditions.

Upregulated		
CDS	**Annotation**	**Fold change**
Traes_3B_DDA4A7673.1	30S ribosomal protein S15	7.37
Traes_1DL_DBAFFE190.1	Cytochrome c biogenesis protein CcsA	6.24
Traes_3B_C7A224494.1	Serine/threonine.protein kinase	4.75
Traes_2AS_86E424632.1	Eukaryotic translation initiation factor 3 subunit H	4.38
Traes_7DS_CCFC4F086.1	Ribonucleoside.diphosphate reductase	3.54
Traes_7AS_B4593FBFF.1	Profilin	3.45
Traes_1BS_7C71DA159.1	DNA topoisomerase	3.35
Traes_5DL_9BF4A2140.2	Photosystem II reaction center protein K	3.30
Traes_3B_2D34DA636.2	Peptidyl.prolyl cis.trans isomerase	3.20
Traes_4DL_496B8485B.2	Proteasome subunit beta type	3.16
Traes_3DL_E2C3A375E.1	30S ribosomal protein S4, chloroplastic\x3b Rps4\x3b Rps4 protein	3.16
Traes_7BL_8C6B85F0D1.1	Histone H2A	3.08
Traes_4DL_ABAB935A7.1	50S ribosomal protein L22, chloroplastic	3.04
Traes_4BL_47A020E26.1	BEL1.type homeodomain protein\x3b Uncharacterized protein	3.02
Traes_5BL_8112F0DD3.1	Ubiquitin.conjugating enzyme E2.23 kDa	2.99
Traes_2BL_C433BB333.3	Cytochrome b6.f complex subunit 6	2.89
Traes_1AL_8955C1103.2	60S ribosomal protein L36	2.81
Traes_5AL_15765F705.1	Cytochrome b6	2.80
EPlTAET00000006356	Plant signal recognition particle RNA	2.75
Traes_6BS_9CDECE1A3.1	Histone H2A	2.73
Traes_6BL_D4D9732C8.1	40S ribosomal protein S24	2.71
Traes_7AS_10F2E60A8.1	Cytochrome b6	2.66
Traes_1AL_55440EF31.1	Annexin	2.64
Traes_5DL_BB2DEEC83.1	Histone H3	2.59
Traes_7AL_BE3253C8F.1	ATP.dependent Clp protease proteolytic subunit	2.59
Traes_5BL_56C52F2D9.1	Ribulose bisphosphate carboxylase small chain	2.58
Traes_2AL_4572BEF15.1	Serine/threonine.protein kinase	2.58
Traes_1AL_494A2BAD5.1	Non.lysosomal glucosylceramidase	2.56
Traes_1AL_F4FE96E3B.1	Phenylalanine ammonia.lyase	2.54
Traes_5DL_F8704A24F.2	Serine/threonine.protein kinase	2.48
Traes_2BS_94F8EF65A.1	Histidinol dehydrogenase, chloroplastic	2.46
Traes_1BL_5C78DBF62.1	Malic enzyme	2.44
EPlTAET00000003949	Metazoan signal recognition particle RNA	2.42
EPlTAET00000007936	Plant signal recognition particle RNA	2.42
Traes_4BL_EEB40147B.2	DNA.directed RNA polymerase	2.37
Traes_5AL_0506ABDA3.1	NAD(P)H.quinone oxidoreductase subunit I, chloroplastic	2.36
Traes_1DS_41690F527.1	Plasma membrane ATPase	2.36
Traes_2AS_9AEA9BDEA.1	Serine/threonine.protein kinase	2.34
Traes_6DS_273430303.2	rRNA N.glycosidase	2.33
Traes_2AS_92497D630.1	Glutamate receptor	2.32
Traes_3AL_D8A132EFA.9	Non.specific lipid.transfer protein	2.31

Table 2 (Cont.). Significant differentially expressed genes evaluated in cultivar pioneer 26R61 under waterlogging vs. non-waterlogging conditions.

Upregulated		
CDS	**Annotation**	**Fold change**
Traes_6DL_8C0979724.2	Ribosomal protein L2\x3b Uncharacterized protein	2.27
Traes_6AS_C4E616554.1	GrpE protein homolog	2.27
Traes_4DS_F2FFC249F.1	Histone H2B	2.26
Traes_5AL_96E30FD9B.1	Histone H2B	2.26
Traes_4AL_E6438F411.2	Histone H2A	2.25
Traes_3DL_8C6D663C5.1	Cytochrome c oxidase subunit 3	2.25
Traes_1AL_096D3C715.1	Farnesyl pyrophosphate synthase A2	2.24
Traes_2BL_F4B5C2D79.1	Cytochrome c oxidase subunit 3	2.23
Traes_3AL_5FAA2CCB0.1	DNA.directed RNA polymerase	2.23
Traes_2BS_209788B1B.1	Aldose 1.epimerase	2.22
EPlTAET00000003839	Metazoan signal recognition particle RNA	2.21
Traes_4DL_986C37E09.2	Cytochrome c oxidase subunit 1	2.21
Traes_3AS_2CD6A2085.1	ATP synthase subunit a, chloroplastic	2.19
Traes_3AL_09FA618C6.1	ATP synthase subunit 9, mitochondrial	2.18
EPlTAET00000003933	Plant signal recognition particle RNA	2.16
Traes_4AS_50F2C4B5D.1	30S ribosomal protein S19, chloroplastic	2.13
Traes_7AL_8A07184BC.1	Catalase	2.12
Traes_4DS_DBAA2CC451.2	30S ribosomal protein S11, chloroplastic\x3b DNA.directed RNA polymerase subunit alpha	2.11
Traes_2DL_47ACB537A.1	Auxin response factor	2.11
Traes_3DL_C08B460B8.1	Malic enzyme	2.10
Traes_2DL_1206EF3F9.1	Potassium transporter	2.09
Traes_4BS_1DCF82CB7.1	NAD(P)H.quinone oxidoreductase subunit 5, chloroplastic	2.08
Traes_5AS_2BDDAC590.2	BZip type transcription factor bZIP1\x3b Uncharacterized protein	2.08
Traes_6BL_6977B343C.1	MYB33	2.08
Traes_5AL_A041A47C4.1	Uridine kinase	2.05
Traes_5DL_3675AE907.2	Cytochrome c oxidase subunit 1	2.04
Traes_4AL_DECF2895B.1	Acyl carrier protein	2.03
Traes_7BS_1F4C5C328.1	Profilin	2.03
Traes_4DS_F5C4EA98C.2	NAD(P)H.quinone oxidoreductase subunit K, chloroplastic	2.03
Traes_2AS_B2D2323BC.1	NAD(P)H.quinone oxidoreductase chain 4, chloroplastic	2.02
Traes_4DL_319790D4E.1	3.ketoacyl.CoA synthase	2.01
Downregulated		
CDS	**Annotation**	**Fold change**
Traes_2AS_2FBCA527B.2	Serine/threonine.protein kinase	-7.24
Traes_5DS_3186C8F7D.1	50S ribosomal protein L20, chloroplastic	-5.76
Traes_5DS_7FD4F7284.1	Phosphoglycerate kinase	-5.32
Traes_1BS_71EBCA90B.1	Beta.galactosidase	-4.63
Traes_3AS_373FCB7E9.1	Tubby.like F.box protein	-2.72

Table 2 (Cont.). Significant differentially expressed genes evaluated in cultivar pioneer 26R61 under waterlogging vs. non-waterlogging conditions.

Downregulated		
CDS	**Annotation**	**Fold change**
Traes_6BS_6F8B79003.2	Galactosyltransferase\x3b Uncharacterized protein	-2.46
Traes_6DS_2F77E1434.1	Eukaryotic translation initiation factor 3 subunit A	-2.44
Traes_7DL_BD41F4558.1	Xyloglucan endotransglucosylase/hydrolase	-2.41
Traes_3B_ADCF93AE0.4	Eukaryotic translation initiation factor 3 subunit L Actin.depolymerizing factor 7\x3b	-2.32
Traes_1AL_F1BBE97D5.1	Uncharacterized protein	-2.31
Traes_2AS_352D59E22.1	Protein H2A.7	-2.27
Traes_6BL_5C168B1DD.1	MATE efflux family protein	-2.27
Traes_2DS_2E5286F5D.1	Cytosine.specific methyltransferase	-2.23
Traes_4AL_F86AA1EB8.1	DNA topoisomerase	-2.18
Traes_2AL_5E24D615B.1	Pyruvate kinase	-2.14
Traes_5AL_DCEF8DEFD.1	Mitochondrial Rho GTPase	-2.04

Table 3. CDS used for gene validation and marker sequences used for qRT-PCR.

Candidate Genes	Primer Forward	Primer Reverse
Traes_2AL_6F3A92BF2.1	CCAAGGGTCTCGTTGTGTT	GGCCTTAGTCTCTTCCTTCTTG
Traes_5DS_7FD4F7284.1	CAGGGAGGGCTTTGTAGATATTC	GGTTTCCTTGGTCTCCATCAT
Traes_4AL_CC989B4A5.1	TAGTGCAATTCAGAGCAGAGG	CACCGCAACAACAACAAGAG
Traes_5BL_D526A626E.2	GAACAATCGGACCTGGGAAA	GCTGGTGACTCCTCAACATAC
Traes_5BL_56C52F2D9.1	GCCGTTCATCCTATCACTTCA	CACGATTCCAATTCCAAGTTCTC
Traes_2AL_520618712.1	GACTAGCCGTTCGGTGTATAAAT	AGCTTCTTGCAGACCTTAGC

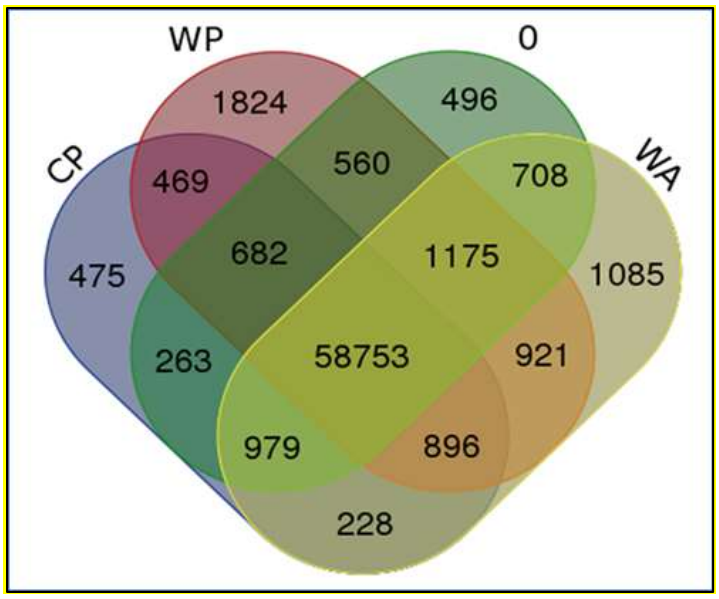

Figure 1. Four way Venn diagram with the distribution of differentially expressed genes across all pairwise comparisons. The number within each shaded area is the number of differentially expressed genes common in each compared treatments; **CP:** control pioneer 26R61; **WP:** waterlogging pioneer 26R61: **CA:** Control AGS2000; **WA:** Waterlogging AGS2000.

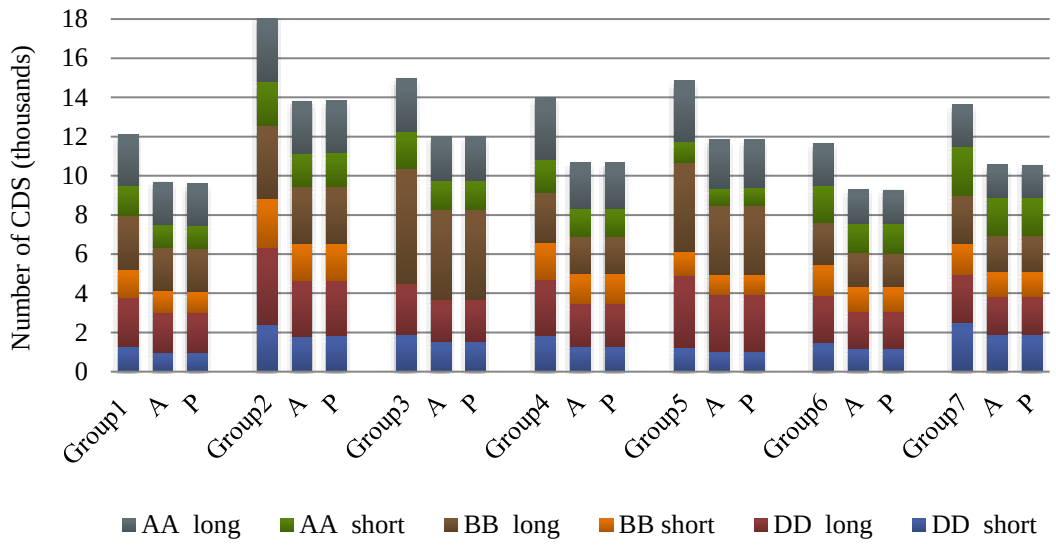

Figure 2. CDS distribution of the seven wheat chromosomes compared with the data base *Triticum aestivum* L. in the data base *EsemblPlants.* **AA:** A genome; **BB:** B genome; **DD**: D genome; **A:** AGS2000 and **P:** Pioneer 26R61.

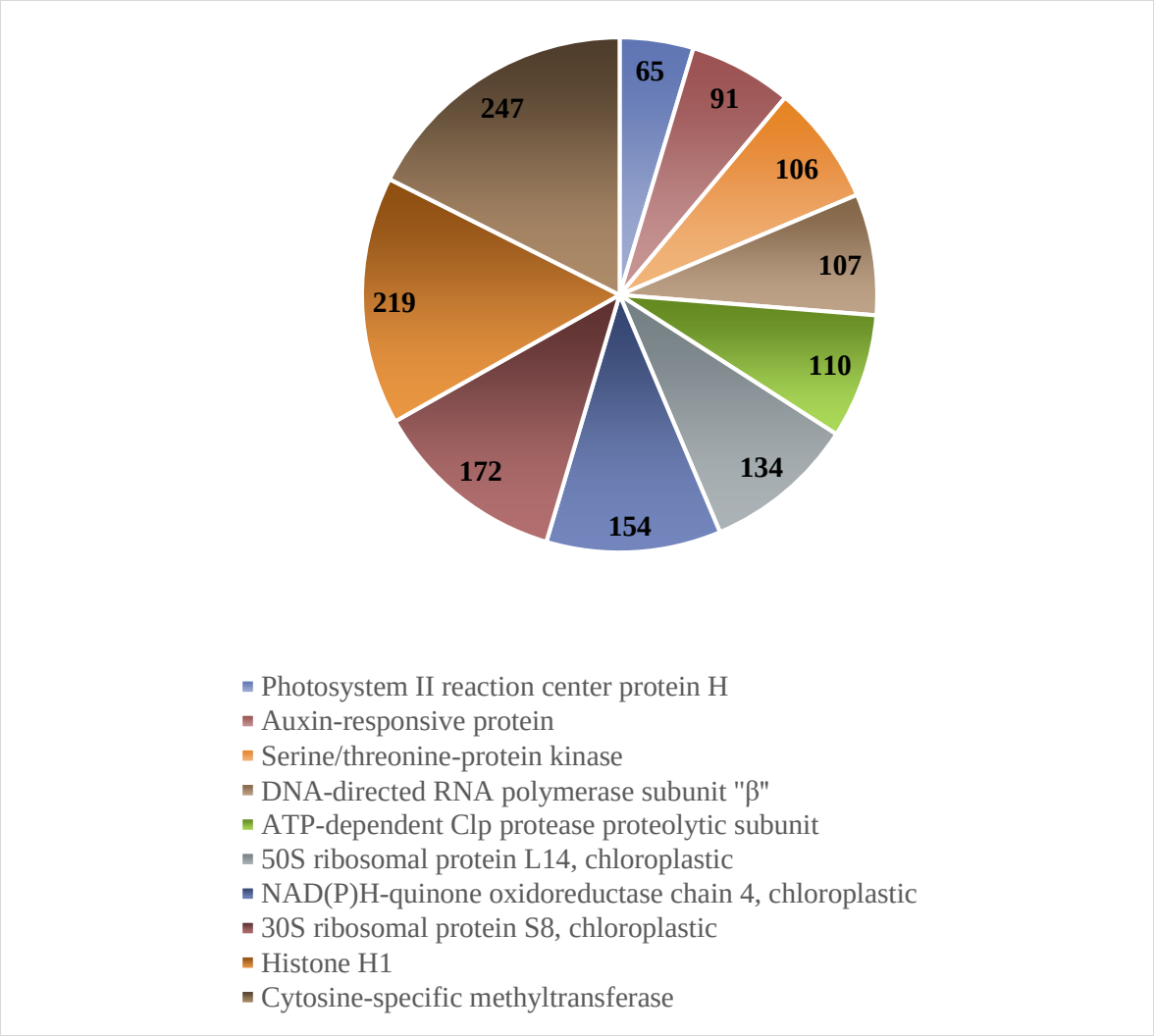

Figure 3. GO result showing the top ten biological processes with the most number of CDS observed under waterlogging conditions and with data from both cultivars AGS2000 and P26R61.

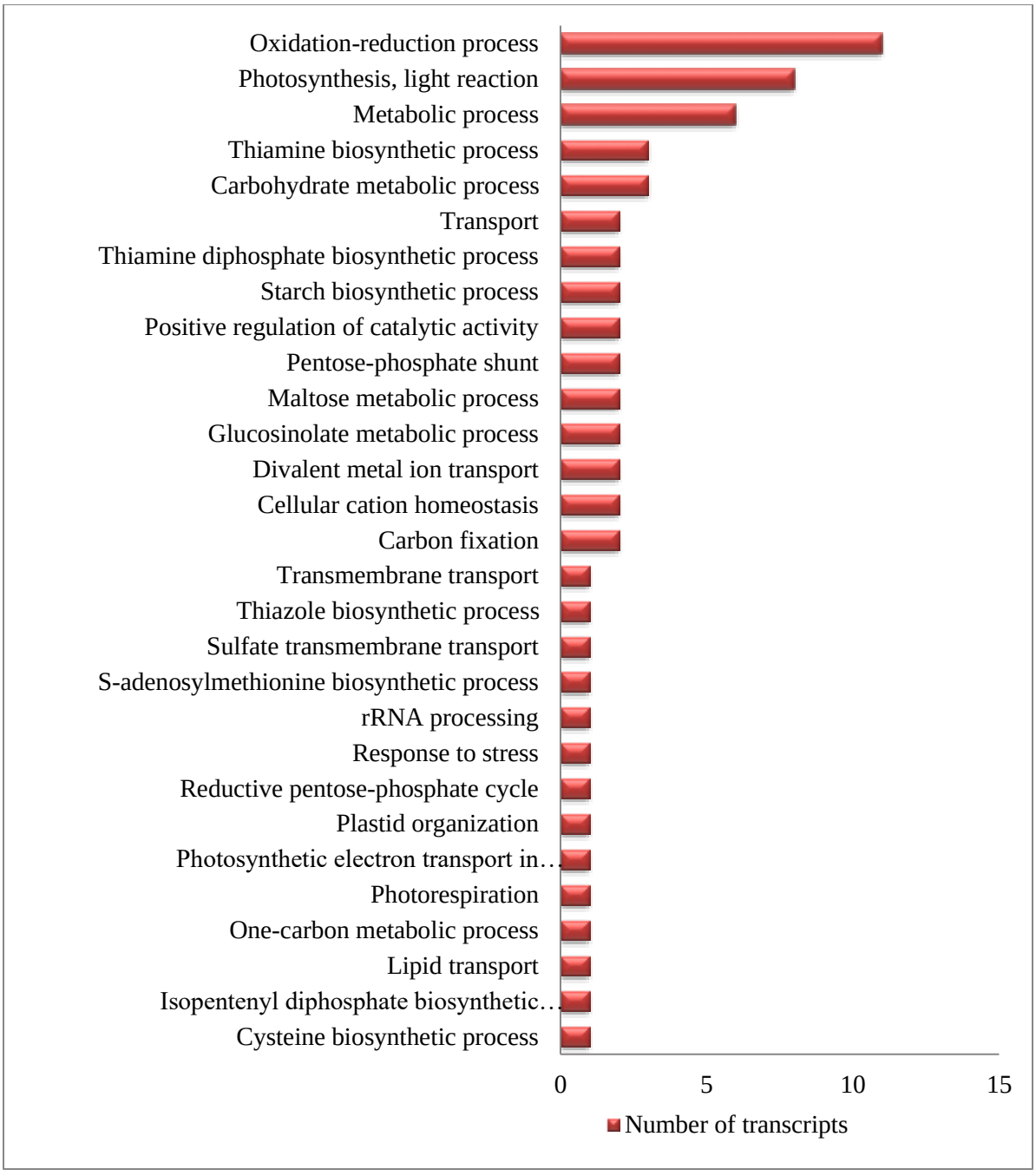

Figure 4. Distribution of gene ontology (GO) biological processes of upregulated CDS in cultivars AGS2000 and pioneer 26R61 under waterlogging conditions.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651-1656

Ahuja. I, De Vos. RCH, Bones. AM, , , Hall. RD (2010) Plant molecular stress responses face climate change. Trends in Plant Science 15:664 - 674

Alam I, Lee DG, Kim KH, Park CH, Sharmin SA, Lee H, Oh KW, Yun BW, Lee BH (2010) Proteome analysis of soybean roots under waterlogging stress at an early vegetative stage. Journal of biosciences 35:49-62

Bailey-Serres J, Fukao T, Gibbs DJ, Holdsworth MJ, Lee SC, Licausi F, Perata P, Voesenek LA, van Dongen JT (2012) Making sense of low oxygen sensing. Trends in plant science 17:129-138

Barakat A, Szick-Miranda K, Chang F, Guyot R, Blanc G, Cooke R, Delseny M, Bailey-Serres J (2001) The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. Plant physiology 127:398-415

Bhardwaj AR, Joshi G, Kukreja B, Malik V, Arora P, Pandey R, Shukla RN, Bankar KG, Katiyar-Agarwal S, Goel S, Jagannath A, Kumar A, Agarwal M (2015) Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop Brassica juncea. BMC plant biology 15:9

Bokhari SA, Wan XY, Yang YW, Zhou L, Tang WL, Liu JY (2007) Proteomic response of rice seedling leaves to elevated CO2 levels. J Proteome Res 6:4624-4633

Bolger ME, Arsova B, Usadel B (2018) Plant genome and transcriptome annotations: from misconceptions to simple solutions. Brief Bioinform 19:437-449

Branco-Price C, Kaiser Kayla A, Jang Charles JH, Larive Cynthia K, Bailey-Serres J (2008) Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. The Plant Journal 56:743-755

Chae L, Sudat S, Dudoit S, Zhu T, Luan S (2009) Diverse transcriptional programs associated with environmental stress and hormones in the Arabidopsis receptor-like kinase gene family. Mol Plant 2:84-107

Chen W, Yao Q, Patil GB, Agarwal G, Deshmukh RK, Lin L, Wang B, Wang Y, Prince SJ, Song L, Xu D, An YC, Valliyodan B, Varshney RK, Nguyen HT (2016) Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. Frontiers in Plant Science 7:1044

Christianson JA, Llewellyn DJ, Dennis ES, Wilson IW (2009) Global gene expression responses to waterlogging in roots and leaves of cotton (*Gossypium hirsutum* L.). Plant and Cell Physiology 51:21-37

Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, Borrill P, Kettleborough G, Heavens D, Chapman H, Lipscombe J, Barker T, Lu FH, McKenzie N, Raats D, Ramirez-Gonzalez RH, Coince A, Peel N, Percival-Alwyn L, Duncan O, Trösch J, Yu G, Bolser DM, Namaati G, Kerhornou A, Spannagl M, Gundlach H, Haberer G, Davey RP, Fosker C, Palma FD, Phillips AL, Millar AH, Kersey PJ, Uauy C, Krasileva KV, Swarbreck D, Bevan MW, Clark MD (2017) An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res 27:885-896

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. Genome biology 17:1-19

Ding J, Huang Z, Zhu M, Li C, Zhu X, Guo W (2018) Does cyclic water stress damage wheat yield more than a single stress? PLoS One 13:e0195535

Du H, Zhu J, Su H, Huang M, Wang H, Ding S, Zhang B, Luo A, Wei S, Tian X, Xu Y (2017) Bulked segregant RNA-seq reveals differential expression and snps of candidate genes associated with waterlogging tolerance in maize. Frontiers in Plant Science 8:1022

Falcone Ferreyra ML, Pezza A, Biarc J, Burlingame AL, Casati P (2010) Plant L10 ribosomal proteins have different roles during development and translation under ultraviolet-B stress. Plant Physiol 153:1878-1894

Fan HH, Wei J, Li TC, Li ZP, Guo N, Cai YP, Lin Y (2013) DNA methylation alterations of upland cotton (Gossypium hirsutum) in response to cold stress. Acta Physiol Plant 35:2445-2453

Finnegan EJ, Kovac KA (2000) Plant DNA methyltransferases. Plant Mol Biol 43:189-201

Gu C, Chen S, Liu Z, Shan H, Luo H, Guan Z, Chen F (2011) Reference gene selection for quantitative real-time PCR in *Chrysanthemum* subjected to biotic and abiotic stress. Mol Biotechnol 49:192-197

Gupta KJ, Zabalza A, van Dongen JT (2009) Regulation of respiration when the oxygen availability changes. Physiol Plant 137:383-391

Hattori Y, Nagai K, Furukawa S, Song XJ, Kawano R, Sakakibara H, Wu J, Matsumoto T, Yoshimura A, Kitano H, Matsuoka M, Mori H, Ashikari M (2009) The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature 460:1026-1030

Herzog M, Striker GG, Colmer TD, Pedersen O (2016) Mechanisms of waterlogging tolerance in wheat--a review of root and shoot physiology. Plant, cell & environment 39:1068-1086

Hincha DK, Meins Jr F, Schmitt JM (1997) [beta]-1,3-Glucanase is cryoprotective in vitro and is accumulated in leaves during cold acclimation. Plant Physiol 114:1077-1083

Klok EJ, Wilson IW, Wilson D, Chapman SC, Ewing RM, Somerville SC, Peacock WJ, Dolferus R, Dennis ES (2002) Expression profile analysis of the low-oxygen response in Arabidopsis root cultures. Plant Cell 14:2481-2494

Kosová K, Vítámvás P, Prášil IT, Renaut J (2011) Plant proteome changes under abiotic stress Contribution of proteomics studies to understanding plant stress response. Journal of Proteomics 74:1301-1322

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359

Lasanthi-Kudahettige R, Magneschi L, Loreti E, Gonzali S, Licausi F, Novi G, Beretta O, Vitulli F, Alpi A, Perata P (2007) Transcript profiling of the anoxic rice coleoptile. Plant Physiology 144:218-231

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Liu F, Zhang X, Lu C, Zeng X, Li Y, Fu D, Wu G (2015) Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. Journal of experimental botany 66:5663-5681

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25:402-408

MacKintosh C, Lyon GD, MacKintosh RW (1994) Protein phosphatase inhibitors activate anti-fungal defence responses of soybean cotyledons and cell cultures. The Plant Journal 5:137-147

Mahajan S, Tuteja N (2005) Cold, salinity and drought stresses: An overview. Archives of Biochemistry and Biophysics 444:139 - 158

Moore G, Devos K, Wang Z, Gale M (1995) Cereal genome evolution: grasses, line up and form a circle. Current biology 5:737-739

Oh MW, Nanjo Y, Komatsu S (2014a) Gel-free proteomic analysis of soybean root proteins affected by calcium under flooding stress. Front Plant Sci 5

Oh MW, Nanjo Y, Komatsu S (2014b) Identification of nuclear proteins in soybean under flooding stress using proteomic technique. Protein and peptide letters 21:458-467

Ponnamperuma FN (1972) The chemistry of subemerged soils. Advances in agronomy 24:29-96.

Qi X-H, Xu X-W, Lin X-J, Zhang W-J, Chen X-H (2012) Identification of differentially expressed genes in cucumber (*Cucumis sativus* L.) root under waterlogging stress by digital gene expression profile. Genomics 99:160-168

Rita S, K. MSR, Garima M, Priyanka D, K. TA, Sanjay K, Meenu K (2009) Rice cytosine DNA methyltransferases – gene expression profiling during reproductive development and abiotic stress. The FEBS Journal 276:6301-6311

Sage RF, Sharkey TD, Seemann JR (1988) The in-vivo response of the ribulose-1, 5-bisphosphate carboxylase activation state and the pool sizes of photosynthetic metabolites to elevated CO2 inPhaseolus vulgaris L. Planta 174:407-416

Saha A, Das S, Moin M, Dutta M, Bakshi A, Madhav MS, Kirti PB (2017) Genome-wide identification and comprehensive expression profiling of ribosomal protein small subunit (RPS) genes and their comparative analysis with the large subunit (RPL) genes in rice. Frontiers in Plant Science 8:1553

Setter TL, Waters I, Sharma SK, Singh KN, Kulshreshtha N, Yaduvanshi NP, Ram PC, Singh BN, Rane J, McDonald G, Khabaz-Saberi H, Biddulph TB, Wilson R, Barclay I, McLean R, Cakir M (2009) Review of wheat improvement for waterlogging tolerance in Australia and India: the importance of anaerobiosis and element toxicities associated with different soils. Ann Bot 103:221-235

Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J Biol Chem 277:37741-37746

Tamang B, Fukao T (2015) Plant adaptation to multiple stresses during submergence and following desubmergence. International journal of molecular sciences 16:26226

Tenea GN, Peres Bota A, Cordeiro Raposo F, Maquet A (2011) Reference genes for gene expression studies in wheat flag leaves grown under different farming conditions. BMC research notes 4:373

Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. The Plant journal : for cell and molecular biology 57:1065-1078

Vazquez-Tello A, Ouellet F, Sarhan F (1998) Low temperature-stimulated phosphorylation regulates the binding of nuclear factors to the promoter of Wcs120, a cold-specific gene in wheat. Mol Gen Genet 257:157-166

Wada Y (2005) Physiological functions of plant DNA methyltransferases. Plant biotechnology 22:71-80

Whittle CA, Krochko JE (2009) Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in Brassica napus. Plant Cell 21:2203-2219

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail A, Bailey-Serres J, Ronald P, Mackill D (2006) *Sub1A* is an ethylene responsive-factor-like gene that confers submergence tolerance to rice. Nature 442:705 - 708

Yasuo T, Tadashiro K, Takao K (1956) Microbial metabolism in reduction process of paddy soils (Part 1). Soil Science and Plant Nutrition 2:63-66

Yu RM, Zhou Y, Xu ZF, Chye ML, Kong RY (2003) Two genes encoding protein phosphatase 2A catalytic subunits are differentially expressed in rice. Plant Mol Biol 51:295-311

CONCLUSIONS

Over-all conclusions

A cross-validation approach was performed to evaluate micro and macronutrient concentrations in 240 winter wheat lines to determine model accuracy. Moderate to high prediction accuracies were obtained for Ca, K, Mg, S, Mn, Na and Zn indicating that the model developed could be implemented in the Wheat Breeding Program at the University of Arkansas to identify tolerant lines without costly and laborious phenotypic evaluation. Additionally, a genome wide association analysis was performed identifying genomic regions associated with elemental concentrations. Aluminum and iron concentrations were shown to be positive correlated in all environments and negative correlated with kernel weight per spike. Therefore, multi-trait markers identified in this study are an important resource for line selection for different elemental toxicities for wheat improvement under waterlogging conditions. Finally, evaluation of gene expression was performed in two winter wheat cultivars using RNA-seq analysis. Identification of CDS associated with previously reported molecular and biological processes associated with abiotic stresses are reported. Gene annotation was shown to be an important limiting aspect for gene identification in this study. However, being one of the first studies evaluating the global wheat transcriptome under waterlogging conditions the CDS here identified could be used as candidate genes for wheat waterlogging stress. The three methods used here provide new information to understand different aspects of soil waterlogging stress and waterlogging stress tolerance and their application to wheat improvement.

APPENDICES

Supplementary 1. Soil analyses results from Stuttgart, AR site using the Spectro Arcos ICP_OES analyzer.

Site-Year	Treatment	pH	EC	P	K	Ca	Mg	S	Na	Fe	Mn	Zn	Cu	B
	Pre-WLST	5.51	77.00	24.75	160.00	1193.00	125.50	12.35	55.75	464.50	123.00	2.75	1.00	0.30
ST13	Non-WL	5.49	109.50	20.60	130.87	1252.96	121.74	10.75	64.82	463.94	124.86	2.49	0.98	0.29
	WL	5.30	131.50	24.77	161.90	1200.88	130.83	12.16	60.72	462.46	124.28	2.75	1.03	0.29
ST14	Non-WL	5.22	91.00	21.15	97.84	1106.35	106.72	5.35	42.30	533.98	60.23	0.59	1.33	0.30
	WL	5.30	142.00	21.81	142.24	1151.00	123.81	7.89	54.32	495.03	160.49	2.74	1.16	0.28

Supplementary 2. Highly significant ($p < 0.00001$), very significant and consistent ($p < 0.0005$) and significant and consistent ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

Highly significant < 0.00001

SNP	Trait	Env	Chr	Position (bp)	P. value	MAF	R^2 [a]	Allelic Effect [b]
S1B_599416338	Na	SJ14	1B	599416338	1.5E-08	0.22	0.23	20.45
S1B_599652508	Zn	ST13	1B	599652508	4.9E-07	0.05	0.16	3.66
S1B_599652549	Zn	ST13	1B	599652549	4.9E-07	0.05	0.16	3.66
S1B_600641545	Na	SJ14	1B	600641545	7.7E-07	0.24	0.20	16.86
S1B_599772225	Na	SJ14	1B	599772225	6.8E-08	0.23	0.22	19.02
S1B_598061897	Na	SJ14	1B	598061897	7.9E-08	0.23	0.22	19.08
S1B_598092427	Na	SJ14	1B	598092427	6.8E-08	0.22	0.22	19.38
S1B_599772202	Na	SJ14	1B	599772202	5.4E-08	0.23	0.22	19.47
S1B_597999814	Na	SJ14	1B	597999814	6.0E-08	0.22	0.22	19.61
S1B_599660679	Na	SJ14	1B	599660679	3.3E-08	0.22	0.22	19.68
S1B_597722660	Na	SJ14	1B	597722660	2.6E-08	0.22	0.23	20.15
S1B_599660677	Na	SJ14	1B	599660677	2.6E-08	0.22	0.23	20.15
S1B_599652390	Na	SJ14	1B	599652390	2.1E-08	0.23	0.23	20.15
S1B_599759697	Na	SJ14	1B	599759697	1.5E-08	0.22	0.23	20.42
S1B_599772225	Na	ST13	1B	599772225	6.9E-07	0.23	0.20	175.49
S1B_599652390	Na	ST13	1B	599652390	8.3E-07	0.23	0.20	177.09
S1B_599772202	Na	ST13	1B	599772202	7.7E-07	0.23	0.20	177.13
S1B_599759697	Na	ST13	1B	599759697	8.6E-07	0.22	0.20	177.29
S1B_597999814	Na	ST13	1B	597999814	9.7E-07	0.22	0.20	177.65
S1B_599416338	Na	ST13	1B	599416338	3.6E-07	0.22	0.20	183.89
S1D_41196155	Fe	ST13	1D	41196155	9.9E-07	0.06	0.12	152.54
S7D_610181301	Fe	SJ13	7D	610181301	1.1E-06	0.06	0.17	59.78
S1B_599660679	Na	ST13	1B	599660679	1.2E-06	0.22	0.19	172.93
S7B_117334234	B	SJ14	7B	117334234	1.3E-06	0.06	0.19	0.61
S7B_117334235	B	SJ14	7B	117334235	1.3E-06	0.06	0.19	0.61
S1B_597722660	Na	ST13	1B	597722660	1.4E-06	0.22	0.19	174.43
S1B_599660677	Na	ST13	1B	599660677	1.4E-06	0.22	0.19	174.43
S1A_192014751	Fe	ST13	1A	192014751	1.4E-06	0.07	0.11	153.44
S1B_598061897	Na	ST13	1B	598061897	1.6E-06	0.23	0.19	171.02
S1B_600641545	Na	ST13	1B	600641545	1.6E-06	0.24	0.19	165.04
S1B_46140066	Fe	SJ13	1B	46140066	1.9E-06	0.11	0.11	119.49
S1B_46140047	Fe	ST13	1B	46140047	2.0E-06	0.10	0.11	121.74
S2A_612152711	Na	ST13	2A	612152711	2.2E-06	0.05	0.19	315.79
S2A_612152741	Na	ST13	2A	612152741	2.2E-06	0.05	0.19	315.79
S7D_610181301	Al	SJ13	7D	610181301	2.5E-06	0.06	0.15	34.79
S6B_111561002	Fe	ST13	6B	111561002	2.9E-06	0.08	0.11	131.96

Supplementary 2 (Cont.). Highly significant ($p < 0.00001$), very significant and consistent ($p < 0.0005$) and significant and consistent ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

Highly significant < 0.00001

SNP	Trait	Env	Chr	Position (bp)	P. value	MAF	R^2 [a]	Allelic Effect [b]
S2B_77204516	S	SJ14	2B	77204516	3.7E-06	0.34	0.22	0.01
S1A_584433749	Zn	SJ14	1A	584433749	4.3E-06	0.11	0.22	1.28
S1B_598092427	Na	ST13	1B	598092427	4.6E-06	0.22	0.18	164.66
S3A_702519039	B	SJ13	3A	702519039	6.2E-06	0.11	0.14	0.22
S3B_546349156	P	SJ14	3B	546349156	7.9E-06	0.14	0.19	0.02
SUN_244556964	B	SJ14	NA	244556964	8.0E-06	0.08	0.18	0.45
S1A_176108383	Na	SJ13	1A	176108383	8.1E-06	0.06	0.15	42.53
S1A_176108439	Na	SJ13	1A	176108439	8.1E-06	0.06	0.15	42.53
S7B_568062247	Zn	ST13	7B	568062247	9.1E-06	0.11	0.14	2.05
S4D_238865033	Cu	ST13	4D	238865033	9.9E-06	0.06	0.16	2.36
S4D_238865041	Cu	ST13	4D	238865041	9.9E-06	0.06	0.16	2.36

Very significant and consistent p < 0.0005

SNP	Trait	Env	Chr	Position	P. value	MAF	R^2 [a]	Allelic Effect [b]
S6B_309688673	Ca	SJ13	6B	309688673	3.2E-05	0.07	0.23	-0.03
		SJ14	6B	309688673	1.7E-04	0.06	0.11	-0.03
S6B_309688674	Ca	SJ13	6B	309688674	3.2E-05	0.07	0.23	-0.03
		SJ14	6B	309688674	1.7E-04	0.06	0.11	-0.03
S7D_579146731	Na	SJ13	7D	579146731	5.8E-05	0.06	0.13	38.33
		SJ14	7D	579146731	2.3E-04	0.06	0.15	22.84
S1A_379438389	Fe	SJ14	1A	379438389	4.5E-04	0.40	0.08	11.34
		ST13	1A	379438389	2.6E-04	0.40	0.10	-0.01
S1A_372238511	Fe	SJ14	1A	372238511	3.3E-04	0.41	0.08	11.65
		ST13	1A	372238511	2.2E-04	0.41	0.10	-0.01
S1A_376790586	Fe	SJ14	1A	376790586	2.5E-04	0.40	0.08	11.86
		ST13	1A	376790586	4.7E-04	0.40	0.10	-0.01
S1A_373360656	Fe	SJ14	1A	373360656	2.1E-04	0.40	0.08	12.01
		ST13	1A	373360656	3.3E-04	0.40	0.10	-0.01
S1A_373812006	Fe	SJ14	1A	373812006	2.0E-04	0.41	0.08	12.17
		ST13	1A	373812006	2.9E-04	0.40	0.10	-0.01
S1A_375844331	Fe	SJ14	1A	375844331	1.5E-04	0.41	0.09	12.27
		ST13	1A	375844331	4.7E-04	0.40	0.10	-0.01
S1A_382337060	Fe	SJ14	1A	382337060	1.1E-04	0.41	0.09	12.57
		ST13	1A	382337060	3.6E-04	0.41	0.10	-0.01
S1A_373846553	Fe	SJ14	1A	373846553	1.1E-04	0.40	0.09	12.55
		ST13	1A	373846553	5.0E-04	0.40	0.10	-0.01

Supplementary 2 (Cont.). Highly significant ($p < 0.00001$), very significant and consistent ($p < 0.0005$) and significant and consistent ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint. Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

Very significant and consistent p < 0.0005

SNP	Trait	Env	Chr	Position (bp)	P. value	MAF	R^2 [a]	Allelic Effect [b]
S1A_380869075	Fe	SJ14	1A	380869075	7.3E-05	0.41	0.09	12.95
		ST13	1A	380869075	4.0E-04	0.41	0.10	-0.01
S1B_599660615	Na	ST13	1B	599660615	1.7E-05	0.18	0.17	165.84
		SJ14	1B	599660615	1.0E-04	0.18	0.16	14.76
S5B_588714935	Na	ST13	5B	588714935	1.5E-04	0.09	0.15	197.80
		SJ14	5B	588714935	3.9E-04	0.09	0.14	18.06
S5B_588714936	Na	ST13	5B	588714936	1.5E-04	0.09	0.15	197.80
		SJ14	5B	588714936	3.9E-04	0.09	0.14	18.06
S7D_528997111	Zn	ST13	7D	528997111	2.2E-05	0.11	0.13	1.97
		SJ14	7D	528997111	1.2E-05	0.11	0.21	1.08
S1B_683422400	Zn	ST13	1B	683422400	1.9E-05	0.11	0.13	1.94
		SJ14	1B	683422400	4.6E-05	0.11	0.20	0.97
S7B_567704027	Zn	ST13	7B	567704027	1.8E-05	0.11	0.13	1.99
		SJ14	7B	567704027	4.7E-04	0.11	0.18	0.86
S7B_568012679	Zn	ST13	7B	568012679	1.8E-05	0.11	0.13	1.99
		SJ14	7B	568012679	4.7E-04	0.11	0.18	0.86
SUN_334466963	Al	SJ14	NA	334466963	3.3E-05	0.05	0.10	13.60
		ST14	NA	334466963	3.4E-05	0.05	0.10	35.90
SUN_334466979	Al	SJ14	NA	334466979	3.3E-05	0.05	0.10	13.60
		ST14	NA	334466979	3.4E-05	0.05	0.10	35.90
SUN_221478096	Al	SJ14	NA	221478096	2.3E-04	0.05	0.08	14.90
		ST14	NA	221478096	1.9E-04	0.05	0.17	214.72
S5A_3619242	Fe	ST14	5A	3619242	3.0E-04	0.10	0.08	21.16
		ST13	5A	3619242	3.6E-04	0.10	0.11	1.64

Significant and consistent p < 0.001

SNP	Trait	Env	Chr	Position	P .value	MAF	R^2 [a]	Allelic Effect [b]
S1A_302553240	Na	SJ14	1A	302553240	9.2E-04	0.08	0.14	18.43
		ST13	1A	302553240	8.1E-04	0.08	0.14	187.63
S3A_720852820	Na	SJ14	3A	720852820	5.3E-04	0.06	0.14	18.87
		ST13	3A	720852820	6.2E-04	0.06	0.14	189.26
S3B_50499819	Mg	SJ13	3B	50499819	9.5E-04	0.06	0.13	0.01
		SJ14	3B	50499819	5.6E-04	0.06	0.09	0.01
S6D_466814084	K	SJ14	6D	466814084	7.7E-04	0.23	0.16	0.10
		ST13	6D	466814084	5.6E-04	0.23	0.11	-0.11
S6D_466814117	K	SJ14	6D	466814117	7.7E-04	0.23	0.16	0.10
		ST13	6D	466814117	5.6E-04	0.23	0.11	-0.11
S7B_564915089	Zn	SJ14	7B	564915089	8.1E-04	0.19	0.18	0.72

Supplementary 2 (Cont.). Highly significant ($p < 0.00001$), very significant and consistent ($p < 0.0005$) and significant and consistent ($p < 0.001$) MTA identified on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	Chr	Position	P .value	MAF	R^2 [a]	Allelic Effect [b]
Significant and consistent p < 0.001								
S7B_564915141	Zn	SJ14	7B	564915141	8.1E-04	0.19	0.18	0.72
		ST13	7B	564915141	5.6E-04	0.19	0.10	1.35
S7B_5745497	P	ST13	7B	5745497	8.8E-04	0.46	0.09	0.01
		ST14	7B	5745497	6.9E-04	0.46	0.11	0.01

[a] Reflect the phenotypic variation explained by the marker, R^2 of the model with SNP calculated in GAPIT package in R

[b] Allelic effects with respect to the minor allele

Supplementary 3. Multi-trait loci marker associations identified at p < 0.0005 on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	P. value	MAF	R^{2} [a]	Allelic Effect [b]
S1B_565301411	Al	SJ13	1.3E-04	0.2	0.12	-17.19
	Fe	SJ13	3.9E-04	0.2	0.12	-26.72
S1D_273497781	P	SJ13	2.4E-04	0.39	0.11	-0.01
	Zn	SJ13	2.9E-04	0.39	0.11	-0.75
S2A_694839900	Al	SJ13	4.7E-04	0.08	0.11	21.79
	Fe	SJ13	2.6E-04	0.08	0.12	37.88
S2A_694839905	Al	SJ13	4.7E-04	0.08	0.11	21.79
	Fe	SJ13	2.6E-04	0.08	0.12	37.88
S2A_694840369	Al	SJ13	1.7E-04	0.08	0.12	23.53
	Fe	SJ13	9.6E-05	0.08	0.13	40.61
S2A_694862042	Al	SJ13	1.7E-04	0.08	0.12	23.53
	Fe	SJ13	9.6E-05	0.08	0.13	40.61
S2A_694912728	Al	SJ13	2.5E-04	0.16	0.12	18.06
	Fe	SJ13	4.2E-04	0.16	0.12	28.87
	Mn	SJ13	1.6E-04	0.16	0.17	2.34
S2A_694955335	Al	SJ13	3.2E-04	0.08	0.11	22.98
	Fe	SJ13	2.1E-04	0.08	0.12	39.29
S2A_694955962	Al	SJ13	3.2E-04	0.08	0.11	22.98
	Fe	SJ13	2.1E-04	0.08	0.12	39.29
S2A_695153479	Al	SJ13	2.5E-04	0.08	0.12	23.04
	Fe	SJ13	1.8E-04	0.08	0.12	39.1
S2A_695161516	Al	SJ13	3.2E-04	0.08	0.11	22.98
	Fe	SJ13	2.1E-04	0.08	0.12	39.29
S2A_757563999	Al	SJ13	3.8E-04	0.16	0.11	16.57
	Fe	SJ13	2.0E-04	0.16	0.12	28.87
S2A_757635358	Al	SJ13	3.0E-04	0.06	0.11	24.62
	Fe	SJ13	1.1E-04	0.06	0.13	43.82
S2A_758357118	Al	SJ13	1.9E-04	0.08	0.12	22.83
	Fe	SJ13	6.2E-05	0.08	0.13	40.51
S2B_748525879	Al	SJ13	2.4E-04	0.06	0.12	31.43
	Fe	SJ13	3.7E-04	0.06	0.12	49.82
S3A_705817053	Fe	SJ13	5.0E-04	0.05	0.11	50.64
	Na	SJ13	4.5E-04	0.05	0.11	38.05
S4A_628813806	Al	SJ13	4.9E-04	0.11	0.11	19.96
	Fe	SJ13	2.0E-04	0.11	0.12	35.53
S4A_669386687	B	SJ13	4.5E-04	0.33	0.11	0.12
	Na	SJ13	1.1E-04	0.33	0.12	19.55

Supplementary 3 (Cont.). Multi-trait loci marker associations identified at p < 0.0005 on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	P. value	MAF	R^{2} [a]	Allelic Effect [b]
S4B_645498754	Al	SJ13	6.8E-05	0.13	0.13	19.95
	Fe	SJ13	2.0E-04	0.13	0.12	30.51
S4B_645498804	Al	SJ13	6.8E-05	0.13	0.13	19.95
	Fe	SJ13	2.0E-04	0.13	0.12	30.51
S4D_367552586	Al	SJ13	3.5E-04	0.11	0.11	20.56
	Fe	SJ13	1.5E-04	0.11	0.12	35.85
S5D_496065369	Al	SJ13	2.4E-04	0.05	0.12	30.11
	Fe	SJ13	4.6E-05	0.05	0.13	54.78
S6D_7288616	Al	SJ13	3.4E-04	0.05	0.11	30.73
	Fe	SJ13	1.1E-04	0.05	0.13	54.7
S6D_7288623	Al	SJ13	3.4E-04	0.05	0.11	30.73
	Fe	SJ13	1.1E-04	0.05	0.13	54.7
S7A_502664126	B	SJ13	3.6E-04	0.16	0.11	0.15
	Na	SJ13	2.6E-04	0.16	0.12	22.4
S7A_55552029	B	SJ13	4.8E-05	0.17	0.13	0.17
	Na	SJ13	3.6E-05	0.17	0.13	24.73
S7D_534248332	B	SJ13	2.1E-04	0.05	0.11	0.3
	Ca	SJ13	4.0E-04	0.05	0.21	0.03
S7D_534248350	B	SJ13	2.1E-04	0.05	0.11	0.3
	Ca	SJ13	4.0E-04	0.05	0.21	0.03
SUN_296404812	Al	SJ13	8.4E-05	0.07	0.12	28.68
	Fe	SJ13	2.5E-04	0.07	0.12	43.6
S1A_459729941	Ca	SJ14	8.1E-05	0.21	0.12	0.02
	Mg	SJ14	1.6E-04	0.21	0.1	0
S1A_587879436	B	SJ14	3.0E-05	0.09	0.17	0.4
	Fe	SJ14	3.1E-04	0.09	0.08	20.48
S1D_64688495	S	SJ14	1.8E-04	0.05	0.19	0.02
	Mg	SJ14	1.4E-04	0.05	0.1	0.01
S2A_253056018	S	SJ14	1.1E-05	0.19	0.22	0.01
	Mg	SJ14	4.1E-04	0.19	0.09	0
S2A_461478424	Al	SJ14	1.0E-04	0.07	0.09	12.02
	Fe	SJ14	2.1E-06	0.07	0.12	32.56
S2A_562476935	S	SJ14	3.0E-04	0.46	0.19	0.01
	Mg	SJ14	4.7E-04	0.46	0.09	0
S2A_562476962	S	SJ14	3.0E-04	0.46	0.19	0.01
	Mg	SJ14	4.7E-04	0.46	0.09	0
S2B_287933173	S	SJ14	2.2E-04	0.05	0.19	0.02
	Mg	SJ14	1.6E-04	0.05	0.1	0.01

Supplementary 3 (Cont.). Multi-trait loci marker associations identified at p < 0.0005 on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	P. value	MAF	R^{2} [a]	Allelic Effect [b]
S2B_77204516	Zn	SJ14	1.5E-05	0.34	0.21	0.92
	Mg	SJ14	4.2E-05	0.34	0.18	0.14
S2B_77846905	S	SJ14	1.6E-04	0.32	0.19	0.01
	Zn	SJ14	5.2E-05	0.32	0.2	0.9
S2B_77846934	S	SJ14	2.1E-05	0.33	0.21	0.01
	Zn	SJ14	1.0E-05	0.33	0.21	0.97
	K	SJ14	4.1E-05	0.33	0.18	0.14
S2B_77846935	S	SJ14	2.1E-05	0.33	0.21	0.01
	Zn	SJ14	1.0E-05	0.33	0.21	0.97
	K	SJ14	4.1E-05	0.33	0.18	0.14
S2D_29746125	Zn	SJ14	3.0E-04	0.06	0.19	1.08
	K	SJ14	2.8E-04	0.06	0.17	0.17
S3B_546349131	K	SJ14	3.4E-05	0.05	0.18	0.21
	P	SJ14	3.3E-04	0.05	0.16	0.02
S3B_546349136	K	SJ14	3.4E-05	0.05	0.18	0.21
	P	SJ14	3.3E-04	0.05	0.16	0.02
S4B_649666144	Al	SJ14	3.4E-04	0.08	0.08	8.79
	Fe	SJ14	1.1E-04	0.08	0.09	22.02
S4B_649679016	Al	SJ14	3.6E-04	0.08	0.08	8.75
	Fe	SJ14	1.1E-04	0.08	0.09	22.04
S4B_649679022	Al	SJ14	3.6E-04	0.08	0.08	8.75
	Fe	SJ14	1.1E-04	0.08	0.09	22.04
S4B_653384409	Fe	SJ14	1.4E-04	0.05	0.09	27.11
	K	SJ14	2.2E-04	0.05	0.17	0.22
S4B_654694163	Fe	SJ14	1.4E-04	0.05	0.09	27.11
	K	SJ14	2.2E-04	0.05	0.17	0.22
S4B_655479618	Fe	SJ14	1.4E-04	0.05	0.09	27.11
	K	SJ14	2.2E-04	0.05	0.17	0.22
S4B_656490823	Fe	SJ14	1.4E-04	0.05	0.09	27.11
	K	SJ14	2.2E-04	0.05	0.17	0.22
S5B_693458300	Na	SJ14	3.7E-04	0.06	0.15	21.03
	Fe	SJ14	3.4E-04	0.06	0.08	22.53
S6B_22497451	P	SJ14	4.8E-04	0.29	0.15	0.01
	Zn	SJ14	1.5E-04	0.29	0.19	0.7
S6B_465725766	Zn	SJ14	3.9E-04	0.14	0.18	0.8
	K	SJ14	3.2E-05	0.14	0.18	0.15
S6B_465725783	Zn	SJ14	3.9E-04	0.14	0.18	0.8
	K	SJ14	3.2E-05	0.14	0.18	0.15

Supplementary 3 (Cont.). Multi-trait loci marker associations identified at p < 0.0005 on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	P. value	MAF	R^{2} [a]	Allelic Effect [b]
S7A_164226930	S	SJ14	3.4E-04	0.1	0.19	0.15
	Mg	SJ14	4.2E-05	0.1	0.11	0.01
S7A_29829940	Ca	SJ14	4.6E-04	0.16	0.1	0.02
	Mg	SJ14	2.4E-04	0.16	0.09	0
S7A_677922775	Mg	SJ14	4.2E-04	0.17	0.09	0
	Na	SJ14	4.0E-04	0.17	0.14	13.34
S7B_677620849	P	SJ14	4.2E-04	0.06	0.16	0.02
	Mg	SJ14	3.7E-04	0.06	0.09	0.01
S7D_350307004	S	SJ14	1.4E-04	0.05	0.2	0.02
	Zn	SJ14	2.3E-04	0.05	0.19	1.11
S1A_60402263	Fe	ST13	6.9E-05	0.1	0.08	97.1
	S	ST13	1.2E-04	0.1	0.09	0.01
S1D_41196155	Mg	ST13	7.3E-05	0.06	0.14	0.01
	Al	ST13	2.2E-04	0.06	0.07	101.93
	S	ST13	8.7E-05	0.06	0.09	0.02
S3A_21509392	B	ST13	4.8E-04	0.06	0.08	0.52
	S	ST13	2.2E-04	0.06	0.08	0.02
S3B_781618097	Al	ST13	2.4E-04	0.12	0.07	88.43
	Fe	ST13	1.9E-05	0.12	0.09	114.7
S6A_5206651	K	ST13	1.9E-04	0.24	0.12	0.12
	S	ST13	3.2E-04	0.24	0.08	0.01
S6B_111561002	Na	ST13	1.1E-04	0.08	0.16	193.47
	Al	ST13	1.8E-04	0.08	0.07	93.86
S7A_411470714	Al	ST13	1.3E-04	0.22	0.08	64.43
	Fe	ST13	3.0E-04	0.22	0.07	66.22
S7B_466001283	Al	ST13	4.7E-04	0.1	0.07	80.74
	Fe	ST13	2.3E-04	0.1	0.07	94.88
S7B_702394186	Al	ST13	1.7E-05	0.06	0.1	126.53
	Fe	ST13	1.8E-04	0.06	0.07	121.48
S4D_288289904	Mg	ST14	5.0E-04	0.13	0.16	0
	S	ST14	3.5E-04	0.13	0.13	0.01
S6A_243725002	K	ST14	7.3E-05	0.06	0.1	0.11
	Zn	ST14	3.2E-05	0.06	0.11	1.59
S6A_243725003	K	ST14	7.3E-05	0.06	0.1	0.11
	Zn	ST14	3.2E-05	0.06	0.11	1.59
S6A_458272442	Zn	ST14	1.3E-04	0.06	0.1	1.47
	Mn	ST14	2.7E-04	0.06	0.15	16.54
S6D_463906257	Mn	ST14	1.6E-05	0.13	0.17	12.2
	P	ST14	2.5E-04	0.13	0.12	0.01

Supplementary 3 (Cont.). Multi-trait loci marker associations identified at p < 0.0005 on the 240 lines of the association mapping panel. Data from Stuttgart, AR 2013 (ST13), AR 2014 (ST14), Saint Joseph, LA 2013 (SJ13) and LA 2014 (SJ14).

SNP	Trait	Env	P. value	MAF	R2 [a]	Allelic Effect [b]
S7A_346387603	Mg	ST14	1.1E-04	0.08	0.17	0
	S	ST14	2.7E-04	0.08	0.14	0.01
S7B_712475364	P	ST14	3.3E-04	0.32	0.12	-0.01
	K	ST14	1.2E-05	0.32	0.11	-0.05

[a] Reflect the phenotypic variation explained by the marker, R2 of the model with SNP calculated in GAPIT package in R

[b] Allelic effects with respect to the minor allele