12-15-2014

Compositional Traits in *Sorghum Bicolor* Characterized by Transcriptome, Ionome and Genome-wide Association Analysis

Nadia Shakoor
University of South Carolina - Columbia

Compositional Traits in *Sorghum Bicolor* Characterized by Transcriptome, Ionome and Genome-wide Association Analysis

by

Nadia Shakoor

Bachelor of Science
University of Illinois Urbana-Champaign, 2005

Master of Science
University of Chicago, 2011

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Biological Sciences

College of Arts and Sciences

University of South Carolina

2014

Accepted by:

Stephen Kresovich, Major Professor

Erin Connolly, Committee Member

Jill Anderson, Committee Member

Robert Friedman, Committee Member

John Steffens, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

ACKNOWLEDGMENTS

I would like to first thank my advisor, Dr. Stephen Kresovich. I'm deeply grateful for his support and flexibility in taking on the supervision of this thesis project under such unusual circumstances. Through your insights, counsel, and greater perspective, you have set a true example of excellence as a researcher, role model and mentor. I would also like to give my thanks to my committee members, including Erin Connolly, John Steffens, Bob Friedman and Jill Anderson. Your discussions and flexibility through this process have been invaluable and are sincerely appreciated. Thank you to the members of the Kresovich lab, particularly Davina Rhodes, Zach Brenton, and Rick Boyles for making my time spent in SC truly enjoyable. I would also like to express my appreciation to the administrative and research personnel at Chromatin Inc., particularly Oswald Crasta and Otto Folkerts, for their support of this dissertation project. I would like to thank Greg Ziegler and Ivan Baxter for their interest and support of the sorghum ionomics project. The experience, technique and knowledge I gained in your lab were instrumental towards the completion of this dissertation. I would like to especially thank my entire family for their support, encouragement and endless well-wishes, particularly my parents, Shamsa and Abdush Shakoor. Finally, I would also like to thank my 'host' family in Columbia, SC, Bruce Kowalski and Kathy Pavelcak for opening their home to me while I carried out my research at the University of South Carolina. I am sincerely grateful for your generosity and genuine southern hospitality.

ABSTRACT

To address the challenge of global mineral malnutrition, current biofortification research in crop plants aims to improve mineral concentration and micronutrient bioavailablity via genetic and traditional breeding methods. Many staple food crops are also used as biofuels, and the chemical and mineral composition of these energy crops directly affect biomass quality and subsequent energy output. Identification of genes and QTL that impact mineral and compositional traits in the grain and biomass of major cereals, including sorghum, is fundamental to developing breeding and selection methods aimed at increasing bioavailable minerals and improving biofuel suitability and seed nutritional quality.

A combinatorial strategy using multiple "omics" methods is an effective approach to understand the molecular genetic systems integral to improving mineral profiles. We can utilize genomics to identify genes, markers and QTL for mineral traits of interest, integrate transcriptomics to evaluate the expression of candidate genes identified in target tissues and genotypes, and carry out ionomic studies to identify and characterize relationships and correlations among and between minerals and to identify common genetic or transcriptomic patterns underlying one or more mineral traits.

Bioinformatics platforms and their associated databases are essential for the integration of these three "omics" approaches. With the goal of advancing sorghum functional genomics, we developed an Affymetrix microarray to

quantify global gene expression and demonstrated its ability to measure gene expression from a series of sorghum lines. The array and the associated gene expression data were developed as a new resource for sorghum crop breeding and genomic discovery. Our expression atlas reports the transcript profiles of 78 sorghum tissues representing shoot, root, leaf, and stem from 6 sorghum lines including grain, sweet, and biomass sorghums.

In addition to our sorghum transcriptomics study, we also integrated genomics and ionomics to identify and characterize the molecular genetic component of mineral compositional traits of interest. Here we describe the successful mineral profiling of a diverse panel of sorghum grain utilized to locate QTL for mineral traits. We ascertained genotypic correlations, between-trait correlations and QTL colocation. We observed QTL colocation in addition to strong phenotypic correlation of elements vital to plant and human nutrition, such as iron and manganese, which suggest that common molecular mechanisms may underlie the uptake and metabolism of these important mineral traits. Combined with high reported heritabilities, these analyses are also promising for the development of breeding strategies aimed at simultaneous improvement of multiple mineral traits in sorghum via marker-assisted selection.

The integration of ionomics and genome-wide association allowed us to identify several germplasm sources, genomic loci and candidate genes as potential biofortification targets for marker-assisted breeding and genomic selection.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF TABLES viii

LIST OF FIGURES x

LIST OF ABBREVIATIONS xii

CHAPTER 1: INTRODUCTION AND BACKGROUND TO RESEARCH SCOPE 1

1.1 BACKGROUND TO RESEARCH SCOPE 2

1.2 TRANSCRIPTOMICS AND MICROARRAY PLATFORMS 3

1.3 IONOMICS AND HIGH-THROUGHPUT PHENOTYPING 4

1.4 GENOMICS AND GENOME-WIDE ASSOCIATION MAPPING 5

1.5 OVERVIEW AND SPECIFIC GOALS 8

CHAPTER 2: A SORGHUM BICOLOR EXPRESSION ATLAS REVEALS DYNAMIC GENOTYPE-SPECIFIC EXPRESSION PROFILES FOR VEGETATIVE TISSUES OF GRAIN, SWEET AND BIOENERGY SORGHUMS 11

2.1 ABSTRACT 12

2.2 BACKGROUND 12

2.3 RESULTS AND DISCUSSION 14

2.4 CONCLUSIONS 31

2.5 METHODS 32

CHAPTER 3: INTEGRATING GENOME WIDE ASSOCIATION MAPPING AND IONOMICS TO IDENTIFY LOCI AFFECTING SEED MINERAL CONCENTRATION IN *SORGHUM BICOLOR* 37

3.1 BACKGROUND / ABSTRACT 38

3.2 Introduction 39
3.3 Results 42
3.4 Discussion 52
3.5 Conclusions 56
3.6 Material and Methods 57
Chapter 4: Conclusion 60
4.1 Concluding Remarks and Areas for Improvement 60
4.2 Future Directions 63
References 68
Appendix A – Supplementary Tables 77
Appendix B – Supplementary Figures 121
Appendix C – Copyright Permission 139

LIST OF TABLES

Table 3.1 Mean, standard deviation, and broad sense heritability of seed mineral concentrations from the Sorghum Association Panel averaged across 4 environments 45

Table A.1 Sorghum samples included in the gene expression atlas 77

Table A.2 Phenotypic characteristics of sorghum genotypes included in gene expression atlas 80

Table A.3 Pearson's correlation coefficient of the biological replicates 81

Table A.4 Expression of sorghum homologs with established patterns of expression in related species 82

Table A.5 Gene Ontology classifications in the biological processes category identified using the AgriGO Singular Enrichment Analysis (SEA) 83

Table A.6 Identification of stably expressed genes (CV </= 15%) 85

Table A.7 List of small RNAs (sorghum, maize and sugarcane) included in microarray design 86

Table A.8 Expression of select sucrose metabolizing enzyme/transporter genes 88

Table A.9 Expression of select phenylpropanoid-monolignol biosynthesis pathway genes 89

Table A.10 SAP Panel and population designation 91

Table A.11 Means and standard deviations of seed mineral concentrations from the SAP and CHP 96

Table A.12 Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels 97

Table A.13 Significant SNPs identified in the 2008 SAP experiments 102

Table A.14 Significant SNPs identified in the 2012 SAP experiments 105

Table A.15 Significant SNPs identified in the 2013 SAP experiments 108

Table A.16 Significant SNPs identified in the SAP experiments based on rank average 111

Table A.17 Significant SNPs identified in the 2010 CHP experiments 114

Table A.18 Shared significant SNPs across SAP experiments 116

Table A.19 Significant SNPs shared in multiple elements 118

Table A.20 Sorghum forage, fuel and 'NREL' NIR trait curves 120

LIST OF FIGURES

Figure 2.1 Pearson's correlation matrix of the whole dataset 19

Figure 2.2 Cluster dendrogram of the whole dataset (78 samples) 20

Figure 2.3 Functional category distribution of tissue-specific transcripts 21

Figure 2.4 Classes sharing similar expression patterns 22

Figure 2.5 Number of tissue-specific genes in across sorghum ideotypes 24

Figure 2.6 Distribution of global gene expression across sorghum tissue types .. 25

Figure 2.7 Hierarchical clustering of samples based on expression of sucrose metabolizing enzymes and sucrose transporter genes 27

Figure 2.8 Number of shared and specific expression profiles of genes expressed in multiple tissue types (Atlas genotype) 28

Figure 2.9 Number of shared and specific expression profiles of genes expressed in multiple stem tissues (Atlas genotype)..... 29

Figure 2.10 Hierarchical clustering of tissues based on expression of phenylpropanoid-monolignol biosynthesis pathway genes 31

Figure 3.1 Box plots depicting the distribution of 20 minerals in four experimental populations. 43

Figure 3.2 Correlation network of seed mineral concentrations using rank average data calculated across replicates from SAP association panels. 46

Figure 3.3 Principal component analysis applied to the rank average seed concentrations for 20 minerals in the SAP lines across experiments. 47

Figure 3.4 Manhattan plot displaying Cd GWAS results..... 50

Figure 3.5 Quantile-quantile (Q–Q) of observed P-values against the expected P-values from the GWAS analysis for mineral concentration 50

Figure 3.6 Evolution of genetic variance at each step of the MLMM 51

Figure 3.7 Allelic effect of the SNP marker for Cd on chromosome 2..... 51

Figure B.1 Correlation of RNA expression between Illumina RNA sequencing and Affymetrix GeneChip microarray platform 121

Figure B.2 Pearson's correlation between expression levels determined by microarray and qRT-PCR..... 122

Figure B.3 Expression dynamics across multiple tissue types detected by microarray and qRT-PCR..... 122

Figure B.4 Number of genes expressed in each of the 78 samples..... 123

Figure B.5 Number of tissue-specific small RNAs across sorghum ideotypes ... 124

Figure B.6 Correlation between log normalized seed weight and elemental concentration in all accessions from SAP and CHP association panels..... 125

Figure B.7 Correlation network of seed mineral concentrations calculated across replicates from SAP and CHP association panels. 126

Figure B.8 Principal component analysis applied to the average seed concentrations for 20 minerals in the SAP and CHP association panels 127

Figure B.9 Manhattan plots for 19 minerals displaying GWAS results 129

List of Abbreviations

Al..... Aluminum

As.....Arsenic

B ..... Boron

Ca..... Calcium

Cd..... Cadmium

CHP ..... Chromatin Association Panel

CMLM.....Compressed mix linear model

Co..... Cobalt

Cu.....Copper

DAP .....Days after planting

ExtBIC .....Extended bayesian information criterion

Fe.....Iron

GBS .....Genotyping by sequencing

GEO .....Gene Expression Omnibus

GWA.....Genome-wide association

GWAS..... Genome-wide association study

ICP-MS ..... Inductively-coupled plasma mass spectroscopy

ICP-OES .....Inductively-coupled optical emission spectrometry

K.....Potassium

mBonf .....Multiple bonferroni criterion

Mg.....Magnesium

MLMM ..... Multi-locus mixed model

Mn ..... Manganese

Mo ..... Molybdenum

Na ..... Sodium

Ni ..... Nickel

NIRS ..... Near-infrared reflectance spectroscopy

P ..... Phosphorus

PCA ..... Principal components analysis

Rb ..... Rubidium

Rep ..... Replicate

RMA ..... Robust multi-array analysis

S ..... Sulfur

SAP ..... Sorghum Association Panel

Se ..... Selenium

Sr ..... Strontium

QTL ..... Quantitative Trait Loci

Zn ..... Zinc

CHAPTER 1

INTRODUCTION

The aim of this thesis project was to utilize and integrate transcriptomics, ionomics and genomics to generate scientific resources for the sorghum research community and address open and relevant questions in the area of compositional trait improvement in *Sorghum bicolor*. In an effort to advance sorghum functional genomics, we constructed a sorghum gene expression atlas and examined gene expression patterns of sorghum vegetative tissues from grain, sweet, and biomass sorghums. We also investigated the genetic architecture of key compositional traits in sorghum by carrying out genome-wide association (GWA) mapping to identify regions of the genome at which genetic variation is associated with grain macronutrients, micronutrients, and elements of interest to plant and animal health.

This chapter discusses the relevance of these projects to the research community and introduces some general aspects of the methods and technology implemented to generate the gene expression atlas. We will discuss the implementation of high-throughput ionomics to quantify the concentration of 20 elements in a large diversity panel of sorghum grain, and we will introduce the genome-wide association (GWA) approach taken to identify genetic loci that correlate with mineral and compositional traits of interest.

1.1 BACKGROUND TO RESEARCH SCOPE

Sorghum [S*orghum bicolor* (L.) Moench] is a staple cereal crop for millions of people in the marginal, semi-arid environments of Africa and South Asia. Its unique and advanced ability to grow in regions of low and variable rainfall highlight its potential to impact agricultural productivity in widespread water-limited environments [1, 2]. Originating and evolving across the diverse environmental landscape of Africa, morphological and physiological adaptation strategies has advanced sorghum as a naturally heat and drought-tolerant warm season C_4 grass that is more efficient at utilizing water, nitrogen and energy resources with respect to other major crops, including maize and wheat [1, 3, 4]. Occupying seven million hectares of farmland, the United States is currently the world's top sorghum producer (8.8 million annual metric tons), followed by India (7.0), Mexico (6.9), and Nigeria (4.8) (http://cgiar.org/sorghum). Cultivated in diverse climates and environmental conditions, the challenges of increasing performance and yield on marginal lands and cooler climates remains at the forefront of sorghum improvement efforts worldwide [5, 6].

Sorghum is globally established as an important source of food, feed, sugar and fiber, and recent interest in bioenergy feedstocks also spotlights sorghum as an attractive alternative for sustainable biofuel production. Framed upon the 2009 sorghum reference genome [7], translational genomic resources have been developed that directly impact research in other closely related feedstock grasses, including switchgrass and *Miscanthus* [8, 9]. Comprehensive understanding of the genetic and molecular mechanisms that regulate metabolite

biosynthesis, transport and storage in these species is essential for the efficient development of biofuel feedstocks.

Sorghum is a staple food and forage crop in the developing world, and this knowledge can also be directed towards investigating the genetic potential for increasing the concentrations of bioavailable minerals in the grain and above-ground biomass. Biofortification, or enhancing the amount of a bioavailable nutrients through genetic selection and plant breeding, is a strategy to combat global mineral malnutrition and is largely dependent upon detecting genes and QTLs that influence mineral traits in the edible portions of crop plants [10, 11].

The work presented here is unique in its identification of expression data and association mapping techniques to identify and characterize genetic elements involved in determining the composition of sorghum grain and biomass. We utilized a combinatorial strategy using multiple omics methods as our approach for understanding the molecular genetic systems integral to characterizing and improving compositional profiles in sorghum.

1.2 TRANSCRIPTOMICS AND MICROARRAY PLATFORMS

Transcriptomics is the study of the transcriptome, and the transcriptome is comprised of all of the RNA molecules produced by the genome, under specific circumstances or in a specific cell. High-throughput methods, such as RNA sequencing or microarray analysis allow for large scale expression profiling or comparison of transcriptomes. Such studies allow for the identification of genes that are differentially expressed in distinct tissue types, or in response to different treatments [12].

The gene expression dataset generated and analyzed in this study was produced using an Affymetrix GeneChip whole genome exon array. This type of array utilizes short oligonucleotides to target gene expression levels from RNA. Specifically designed for this study, the sorghum array was developed with probesets that were complementary to a region, usually at the 3' end of the transcript (due to 5' degradation of mRNA), of each mRNA transcript from *Sorghum bicolor*. To correct for systematic effects observed in raw probe intensities, microarray datasets produced by Affymetrix GeneChip arrays, including the datasets from our study, are preprocessed with RMA (log scale robust multi-array analysis) summarization methods. This standard method for preprocessing integrates calculations for background correction and normalization within and between arrays. The probe intensities are further processed to generate an intensity value for each probeset [13].

1.3 IONOMICS AND HIGH-THROUGHPUT IONOMIC PHENOTYPING

Ionomics is the study of the ionome, and the ionome is defined as the mineral nutrient and trace elemental composition of an organism. Ionomics studies involve the quantitative and simultaneous measurement of an organism's elemental composition, providing a snapshot into the functional state of an organism under different conditions. Changes in the ionome can be driven by genetic and developmental differences as well as by biotic and abiotic factors.

Among the primary methods utilized to measure the plant ionome are inductively-coupled plasma mass spectroscopy (ICP-MS) and inductively-coupled plasma optical emission spectrometry (ICP-OES). We utilized ICP-MS in the present study. It is a more sensitive method that subsequently allows for

smaller sample size. Briefly, the ICP functions to ionize the analyte into atoms, which are then detected by mass spectroscopy. Reference standards are utilized to quantify each element of interest in the sample analyte. ICP-MS analysis time is approximately 1-2 minutes per sample, which allows for a high-throughput processing of hundreds of samples [13].

1.4 GENOMICS AND GENOME-WIDE ASSOCIATION MAPPING

Genomics is the study of the genome and the genome is comprised of all of the genes, regulatory sequences, and noncoding regions of an organism's DNA. With genomics- based approaches, including genome sequencing, QTL and genome-wide association mapping, we are able to analyze the function and structure of the genome and the interaction between loci and alleles within the genome to accelerate gene discovery and functional analyses of genes.

QTL identification by association mapping is achieved by determining a statistical association across allelic variants at SNP markers in genotyped population and the phenotypic measurement of the analyzed trait of interest (e.g. mineral concentration). A strong correlation suggests that there may be a causal relationship between the allelic variation present at the locus and the phenotype [14].

A variety of diversity panels and mapping populations have been utilized in the last decade for association studies in staple crop plants including maize and rice [15, 16]. Further, GWAS on elemental concentration in a mapping populations derived from diverse natural accessions has been successfully carried out for a number of plant species including Arabidopsis, maize and barley[17-19].

There is considerable genetic and phenotypic diversity in sorghum, and GWA mapping is well suited for uncovering the genetic basic for complex traits, including seed mineral accumulation. One of the key strengths of association mapping is that a priori knowledge is not necessary to identify new loci associated with the trait of interest. Further, a GWA mapping population is comprised of accessions that have undergone numerous recombination events, allowing for a narrower mapping interval. However, low power to detect rare alleles and alleles with small effect are distinct limitations of GWA mapping. False positive association of SNPs can also be identified in the GWA analysis due to linkage disequilibrium (e.g. GWA-identified allele A does not affect phenotypic trait of interest, but is in linkage disequilibrium with the causal SNP). False positive SNPs may also arise as an artifact of population structure (e.g. GWA-identified allele A does not affect the trait of interest, but happens to be more common in a related group of lines) [14].

Compared to association mapping, traditional QTL mapping methods (which utilize controlled crosses) are more efficient at uncovering rare and potentially deleterious alleles with a large phenotypic effect. However, such alleles may be irrelevant from an evolutionary perspective, as they are not the cause of the range of phenotypic diversity found in nature. Phenotypic variation in a natural population is a reflection of a particular allele's effects as well as its frequency. GWA analysis methods, including those in this study, combine allelic effects and frequency to identify associations between phenotype and sequence variation across individuals in a population [20].

There are several association models that can be utilized for mapping traits. In the naïve model, the additive effects of the genotypic variance explain the heritable variance in the phenotype of interest. The naïve model does not correct for population structure and is based on the assumption that there is independence among the lines and markers used in the study. Realistically, however, the assumptions of this model cannot be met as most lines and markers utilized in a GWA study are not independent of one another. Diversity panels used in association mapping, including ones developed for sorghum, are a mixture of subpopulations with structured ancestry. Differential selection pressure and reproductive isolation of subpopulations lead to changes in allele frequencies that can confound GWA analysis. The results of the naïve GWA model are broadly assumed to produce high rates of false positive associations [21]. To address these concerns, the naïve GWA model can be modified to include terms for population structure (Q) and kinship (K) that result in SNP effect sizes that are adjusted for both subpopulation and individual relatedness. This model is referred to the Q+K mixed model and is an established method to control for population structure and kinship in GWA analysis [22].

Both the naïve and the mixed model are single-locus approaches, and since traits are not generally controlled by single loci, both of these models are prone to spurious (false positive) associations. Developed by Segura et al., the multi-locus mixed model (MLMM) algorithm utilizes multiple loci in the model, which contribute to a higher detection power and lower potential of false discoveries [23]. Briefly, MLMM is based on EMMA [24] and relies on the iterative use of a simple K, or Q+K, mixed-model algorithm. At each step of the MLMM, the phenotypic variance is divided into genetic, random and explained

variance. The most significant marker is included as a cofactor, and the variance components of the model are recalculated. With each successive iteration, the genetic variance approaches zero, and an optimal model including cofactors that explain the genetic fraction of the phenotypic variance is determined. By regarding significant SNPs as cofactors, spurious or false SNPs that correlate with cofactors included in the model no longer have a high significance value.

1.5 OVERVIEW AND SPECIFIC GOALS

The goal of this dissertation was to investigate the genomic architecture regulating key compositional traits in Sorghum bicolor. We pursued this study in two defined aims:

Research Aim 1: Develop and utilize a whole-transcriptome microarray for the identification and characterization of tissue and genotype-specific expression patterns for all identified *Sorghum bicolor* exons and UTR.

Research Aim 2: Generate and analyze grain ionomic profiles and genetic data for a large panel of diverse sorghum lines, including ~400 genetically diverse lines (SAP) and an additional 150 lines (CHP) representing elite, diverse bioenergy and sugar type sorghums.

For Research Aim 1, we generated a record of gene expression in a set of seven tissues and six diverse sorghum genotypes. In total, we analyzed 78 samples with a combination of four different tissue types (shoot, root, leaf and stem), two dissected stem tissues (pith and rind) and six diverse genotypes, which included 6 public sorghum lines (R159, Atlas, Fremont, PI152611, AR2400, and PI455230) representing grain, sweet, forage, and high biomass sorghum ideotypes.

For Research Aim 2, a total of ~650 sorghum lines, consisting of ~400 lines from the Sorghum Association Panel (SAP) and 150 lines from Chromatin, Inc. (CHP), were used for GWA analysis. The ~400 lines of the SAP were selected for genotypic diversity and phenotypic variation, and the 150 genotypes from Chromatin were selected based on agronomic traits and phenotypic diversity. The SAP lines were previously genotyped using a 'genotyping by sequencing' (GBS) method to identify new and informative SNP markers [25]. In order to obtain comparable genetic data for the complete mapping population, we also carried out the GBS method for the 150 CHP lines.

We phenotyped all 650 lines with field-collected observation as well as chemical and nutrient analysis methods (ICP-MS). We have generated extensive genotypic and phenotypic data on the 'core' 650 lines, allowing for QTL identification by association mapping. Throughout the course of this thesis, we surveyed and identified significant single nucleotide polymorphisms (SNP) in functional genes, both predictive and causative, associated with compositional traits in sorghum by tissue and genotype-specific transcriptional profiling and genome-wide association analysis.

We originally hypothesized, and presently, show in this work that careful examination of natural genetic variation in sorghum allows for the discovery of genes and QTL loci that potentially regulate seed mineral composition. This hypothesis was supported by numerous mapping studies in closely related organisms (e.g. maize, sugarcane, and rice) in which mineral/composition QTL(s) have been identified [26-29].

The thesis continues with the following chapters:

Chapter 2 presents the development and analysis of the sorghum microarray and expression atlas. We collected and isolated RNA from 78 sorghum samples from various developmental stages, genotypes and tissue types. We designed (in collaboration with Chromatin/Affymetrix) a whole-transcriptome exon array for Sorghum bicolor, including all identified exons and putative non-coding RNAs. Data analysis included implementation of standard microarray analysis pipelines in R, principal component analysis, hierarchical clustering and calculation of relative gene expression levels in order to examine biological relatedness and identify expression trends among the samples. The results of this project were published in BMC Plant Biology in Jan, 2014 [30].

Chapter 3 describes the GWA analyses performed for seed mineral concentration. We quantified mineral content (20 elements) in sorghum seeds from 4 replicates of the sorghum diversity panel (n=407) and one replicate of a "validation" diversity panel (n=150) using ICP-MS. We identified genotypic correlations, between-trait correlations and SNP/QTL colocation for mineral traits. We also identified significant SNP markers for 20 mineral traits and several genomic loci and candidate genes are presented as potential biofortification targets for marker assisted breeding and genomic selection. The results and analyses of this project will be submitted for journal publication in November 2014.

Chapter 4 shares the conclusions of Research Aims 1 and 2 and suggests areas for improvement. We discuss future directions, specifically applying seed mineral GWAS methodology to other compositional datasets (e.g. NIR dataset).

CHAPTER 2

A SORGHUM BICOLOR EXPRESSION ATLAS REVEALS DYNAMIC GENOTYPE-SPECIFIC EXPRESSION PROFILES FOR VEGETATIVE TISSUES OF GRAIN, SWEET AND BIOENERGY SORGHUMS

[1]

[1] Shakoor, Nadia, Ramesh Nair, Oswald Crasta, Geoffrey Morris, Alex Feltus, and Stephen Kresovich. "A Sorghum bicolor expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums." *BMC plant biology* 14, no. 1 (2014): 35

2.1 ABSTRACT

Background: Effective improvement in sorghum crop development necessitates a genomics-based approach to identify functional genes and QTLs. Sequenced in 2009, a comprehensive annotation of the sorghum genome and the development of functional genomics resources is key to enable the discovery and deployment of regulatory and metabolic genes and gene networks for crop improvement.

Results: This study utilizes the first commercially available whole-transcriptome sorghum microarray (Sorgh-WTa520972F) to identify tissue and genotype-specific expression patterns for all identified *Sorghum bicolor* exons and UTRs. The genechip contains 1,026,373 probes covering 149,182 exons (27,577 genes) across the *Sorghum bicolor* nuclear, chloroplast, and mitochondrial genomes. Specific probesets were also included for putative non-coding RNAs that may play a role in gene regulation (*e.g.,* microRNAs), and confirmed functional small RNAs in related species (corn and sugarcane) were also included in our array design. We generated expression data for 78 samples with a combination of four different tissue types (shoot, root, leaf and stem), two dissected stem tissues (pith and rind) and six diverse genotypes, which included 6 public sorghum lines (R159, Atlas, Fremont, PI152611, AR2400 and PI455230) representing grain, sweet, forage, and high biomass ideotypes.

2.2 BACKGROUND

Global transcriptome profiling provides a means to access gene networks for the discovery of functional connections between genes, mRNAs and their regulatory proteins, and complex traits expressed through coordinated and dynamic gene networks across different tissues and developmental stages [31].

Over the last decade, microarray-based expression profiling has provided a reliable high-throughput platform for genome-wide analysis of gene expression in many organisms. Microarrays offer substantial advantages for functional genomics, as they are increasingly cost-effective, provide a comparable accuracy of expression profiling to RNA-sequencing, and have been shown to provide comprehensive expression data (up to 90% of the transcriptome) in a given tissue [32]. Well-established microarray data analysis tools are also available for querying, visualizing and analyzing the genomes and predicted genes [33, 34], as well as for analyzing the transcriptome profiling data and integrating with other public datasets [35-38].

To provide insight into the sorghum transcriptome, we generated a record of gene expression in a set of seven tissues and six diverse sorghum genotypes. The choice of samples reflects our aim to develop and enrich the current sorghum transcriptome literature. Previous studies have predominantly focused on reproductive tissues, and the majority of these reports do not represent the complete sorghum transcriptome. Several of these studies have also been limited to the reference genome (BTx623) or Keller, a recently resequenced sweet sorghum variety [39-43].

Comparable whole plant transcriptome maps are available for a number of other model species, including *Arabidopsis thaliana* [44], maize (*Zea mays)* [45], barley (*Hordeum vulgare*) [46], rice (*Oryza sativa*) [47, 48], and soybean (*Glycine max*) [49]. These recent transcriptome surveys were constructed with only one genotype or line/accession for their respective species of interest, whereas the present study aims to highlight the practical importance of examining expression profiles across diverse tissue types, developmental stages, as well as genotypes

in order to accurately target genes and metabolic pathways for the efficient development of improved feedstocks.

Fundamental understanding of sorghum genomics is necessary for improving sorghum for agronomic and compositional traits. Specifically, genotypes with high biomass and increased levels of fermentable stem sugars are ideal for developing feedstocks for the biofuel industry. We developed this genomic resource, the whole-transcriptome array as well as the vegetative transcriptome in diverse genotypes and tissues, in order to facilitate the characterization of molecular networks and regulatory mechanisms governing important metabolic pathways including, but not limited to, cell wall biosynthesis for lignocellulosic biomass as well as synthesis, translocation, and storage of fermentable photosynthates for energy content. The relevance of our dataset is demonstrated by genotype and tissue-specific expression of the phenylpropanoid and lignin biosynthetic pathway genes.

Intended as readily available public resource for functional gene characterization, the transcriptome data presented here is available through NCBI's Gene Expression Omnibus (GEO) under accession number GSE49879, and the Sorghum Genome Array is available through Affymetrix (http://affymetrix.com).

2.3 RESULTS AND DISCUSSION

Generation and quality assessment of data

A whole-transcriptome exon array for *Sorghum bicolor* was custom-designed by Chromatin and Affymetrix: *Sorgh-WTa520972F*. This genechip contains 1,026,373 probes covering 149,182 exons (27,577 genes) across the

Sorghum bicolor genome (10 chromosomes), chloroplast and mitochondria. The sequences used to construct the probesets included all identified *Sorghum bicolor* exons from the Sbi1 assembly (http://phytozome.net). Multiple probes were chosen for each exon, with a minimum of one probe per exon and 25 probes per gene. In addition to standard Affymetrix controls, positive controls in the microarray design included probes for constitutively expressed *Sorghum bicolor* genes (actin, ubiquitin and eIF4a1). Probes for intronic regions of actin and ubiquitin were also utilized to determine background expression levels.

To study the sorghum transcriptome and build a gene expression atlas, we collected 78 diverse samples from various developmental stages and tissue types (Table A1). In order to broadly capture sorghum genetic diversity, we included genotypes representing three major ideotypes, including grain, sweet, and bioenergy sorghums. Our study includes R159, an elite grain sorghum characterized by the valuable agronomic traits of uniform growth and disease resistance [50]. Grain sorghum is cultivated primarily for its high starch content, applications in human/animal health and nutrition, and as biofuel feedstock for ethanol production [5] . We also included two sweet sorghums, Fremont and Atlas, that produce increased biomass and accumulate high levels of fermentable carbohydrates in the stem. Additionally, Fremont is drought resistant and flowers early, while Atlas is less susceptible to lodging (due to a stiff stalk phenotype) and flowers later [51]. We also selected three bioenergy or high biomass lines, PI455230, PI152611, and AR2400 that produce increased levels of cellulosic material and are photoperiod sensitive, which allows the plant to produce higher amounts of vegetative matter under long day conditions (Table

A2). PI152611 is specifically a forage line, a fast-growing, highly digestible grass utilized for livestock feed [5, 50]

The primary goal of this study was to obtain relevant and applicable data for the research community developing sorghum as a global feedstock; this research interest guided our sample selection towards vegetative tissues, with a strong bias for stem tissues. A comprehensive transcriptomic profile of sorghum inflorescence and leaf data was recently made available to the community [40]. We compared the leaf RNA sequencing dataset with the present leaf dataset to demonstrate and confirm that our microarray analysis approach towards transcriptome profiling was appropriate. The Spearman correlation of the transcriptome across technologies is 0.61 (Figure B1), which is consistent with several studies comparing RNA-sequencing and microarray methods for genome-wide transcriptome profiling [52-54]. The present comparison corroborates these studies and demonstrates that the microarray platform for expression profiling correlates well with current sequencing methods. With a common goal of crop improvement, complementary datasets such as these generate a core of information that can be explored for the functional characterization of genes and genetic pathways.

We assessed data quality for hybridization by comparing normalized signals of all probe sets between biological replicates using Pearson's correlation analysis. The biological replicates were highly correlated, with an average Pearson's correlation coefficient of 0.99 (Table A3). The highly reproducible results of the replicate data further validate the quality of the microarray platform and present dataset. Previous studies have consistently established strong correlations between qRT-PCR data and microarray data processed using

robust multi-array analysis (RMA) [55, 56]. However, we also tested a small subset of these genes via qRT-PCR to validate the array-generated expression data and expression patterns across multiple tissue types (Figure B2 and B3).

To further assure data quality, we also examined the general expression patterns of well-characterized genes that have been highlighted for tissue-specific expression in previous studies. In microarray experiments with RNA isolated from shoot tips, we observed high expression levels for homologs of SPATULA, a shoot tip transcription factor that is strongly expressed in shoot tips and young leaf primordia [57]. Similarly, the sorghum homolog for TIP2-3, a root-specific aquaporin gene [58], was also expressed at higher levels in our study using root-isolated RNA (Table A4).

Global gene expression patterns

We detected the expression of 19,354 genes in at least one of the 78 samples, representing 70.2% of all genes on the array (27,577 genes). The number of expressed transcripts detected in the various tissues ranged between 10,850 and 11,587 (representing 56 to 60% of all expressed genes on the array). Expressed genes were determined following established methods [45], and with a conservative and arbitrary expression threshold cutoff of 320 (five times the mean normalized signal from intronic gene probes used as controls), we found that 16.4% of genes on the array were detected in all tissues (4256/27,577) (Figure B4). Gene ontology (GO) annotation analysis of these constitutively expressed genes reveals that most are involved in basic biological processes including development, protein synthesis/modification, and signal transduction (Table A5). Similar to published work in maize, expression of constitutive genes varied

among the samples, with the coefficient of variation (CV) ranging from 5% to 129%. With a CV of 10.4%, we identified a ubiquitin-conjugating enzyme, Sb09g023560, as one of the most stably expressed genes (Table A6). This class of genes was also identified in the maize atlas as the most stably expressed among variable tissues [45].

A diverse range of plant tissues was sampled in this study; however, 29.8% of the probesets were not detected above our designated expression threshold level. Several plausible explanations can account for this incomplete expression coverage, including gene expression from specific tissues and/or developmental stages not included in this study, false positive gene models, and levels of expression below detection threshold limits. Further, the arrays were developed utilizing the BTx623 reference sequence and do not capture polymorphisms, copy number variation and presence-absence variation across all the sampled genotypes.

Transcriptome-based classification of sorghum tissues

A Pearson's distance correlation matrix was constructed to compare and evaluate the transcriptome data from each sample (Figure 2.1). This data shows strong correlations among and within the individual tissue types. The associated dendrogram reveals clustering according to tissue type as well as genotype, highlighting the significance of genotype-specific expression in this study (Figure 2.2).

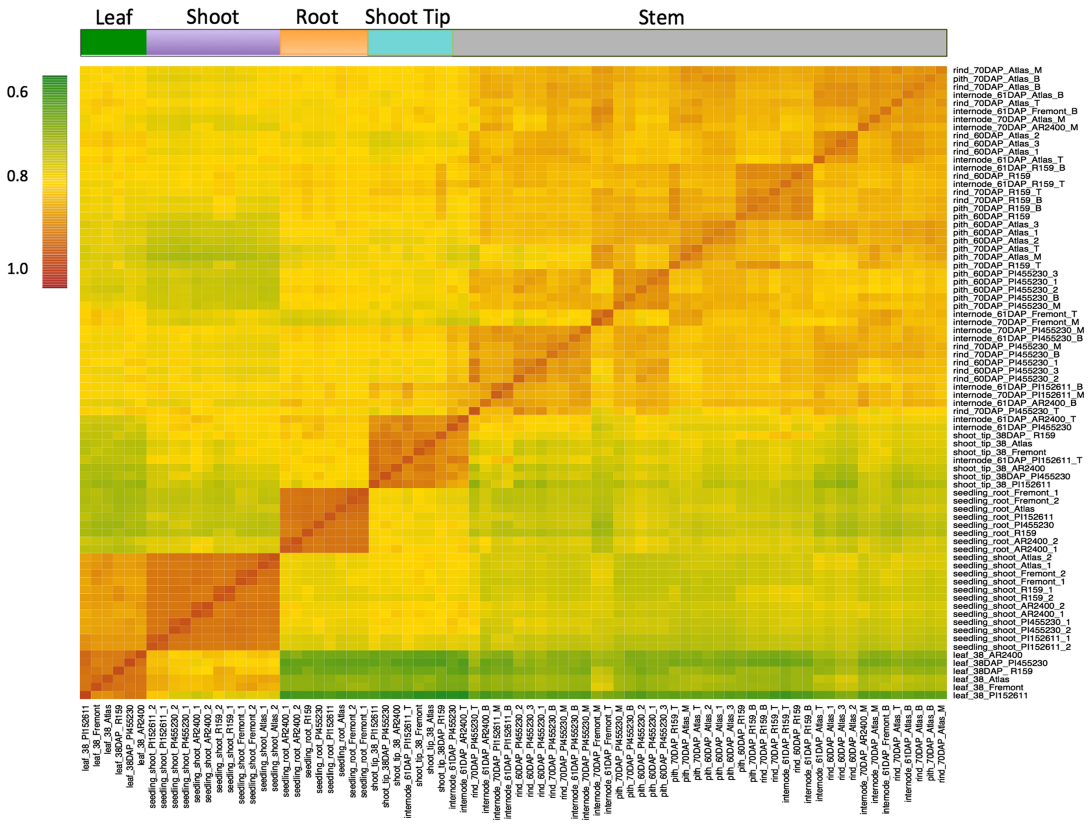


Figure 2.1 Pearson's correlation matrix of the whole dataset. Pair-wise Pearson correlation coefficients were calculated from the gene expression values of the whole transcriptome (27,577 genes) in all 78 samples. The hierarchical clusters were obtained based on Euclidian distance and are indicated by the color bar on the top side of the figure. The color scale indicates the degree of correlation.

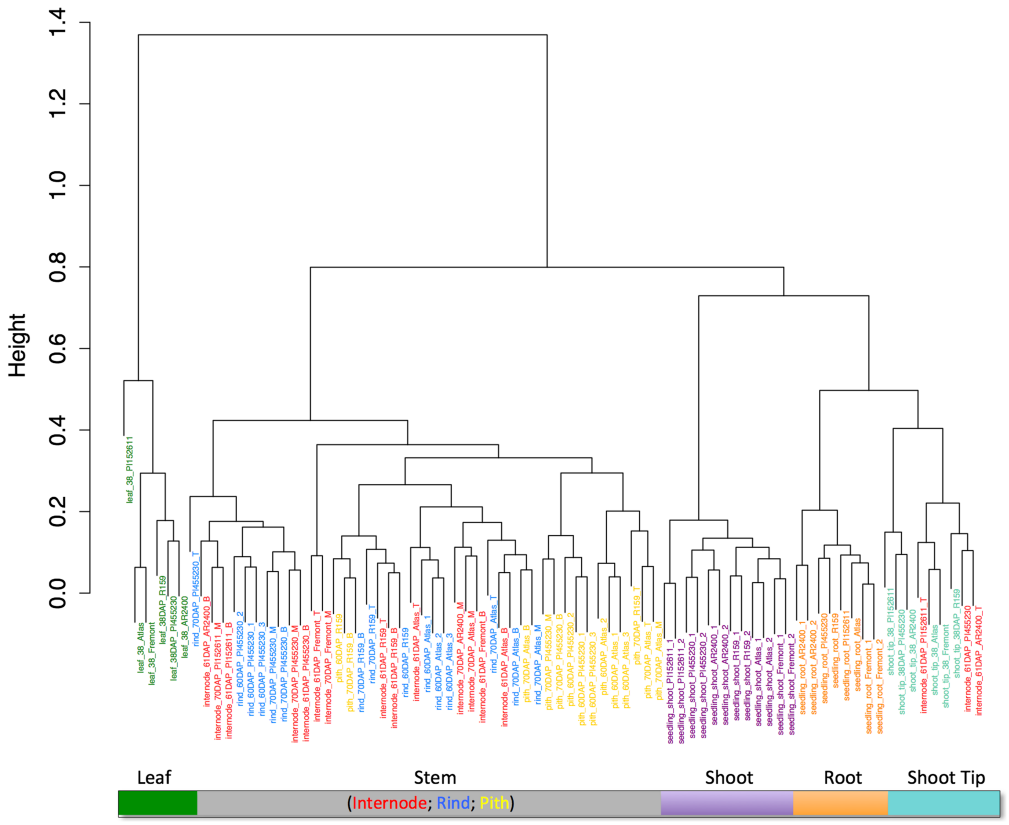


Figure 2.2 Cluster dendrogram of the whole dataset (78 samples). The hierarchical clusters of organs were grouped based on Euclidian distance. The five clusters are indicated by the color bar on the bottom side of the figure.

Utilizing GO categories, functional analysis of the identified gene sets revealed enrichment of known tissue-specific biological processes. For example, the leaf and shoot-associated gene sets were enriched for photosynthetic genes relative to the roots, as expected (Table A5). We found that components of protein synthesis were overexpressed in the seedling roots and shoots, whereas genes involved in metabolism were over-represented in the shoot tip and stem tissues (Figure 2.3). These data identify core sets of genes associated with various biological processes and are clear targets for future study aimed to definitively characterize their functions in specific tissues.

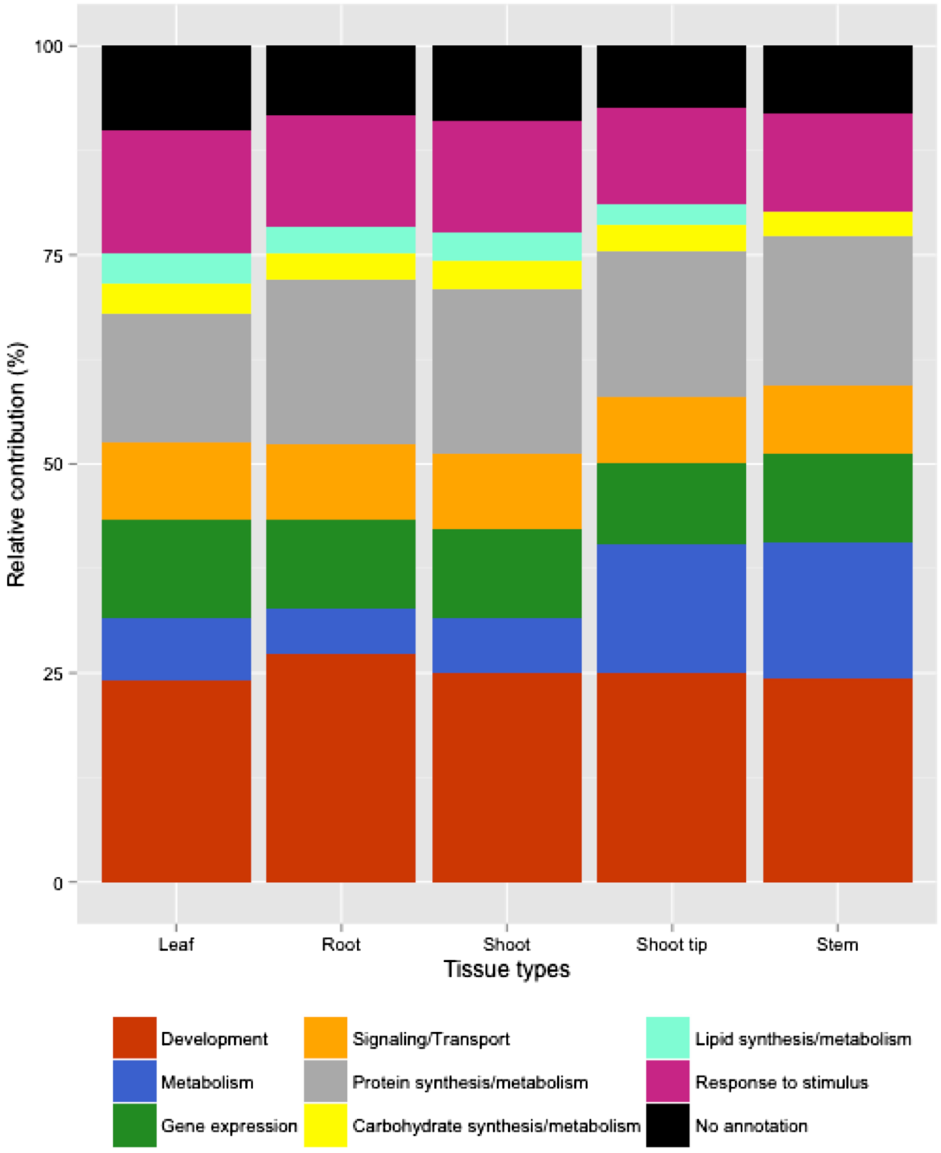


Figure 2.3 Functional category distribution of tissue-specific transcripts. Expression levels of select Gene Ontology categories across tissue types. The Sbi1.4 version of the sorghum annotation allowed for the identification of ~85% of expressed genes across all tissue types. The transcripts were manually verified and grouped into 7 functional categories based on Plant GO slim classifications.

Differential transcriptomes of developmentally distinct vegetative tissues were also apparent from the principal component analysis (PCA) (Figure 2.4). The PCA reveals clustering of functionally related tissue types, and the first two principal components (PC) of this analysis explain 68% of the variance among samples (PC1=48%, PC2=20%). Apical meristematic zones of the roots and shoot tips clustered together and weakly clustered with leaves, shoots and stem tissues. The large group of stem tissues (46 samples) including internode, pith, and rind strongly clustered together and weakly with the remaining tissues. These results are consistent with previous studies in maize and *P. halli* crop models, that show core similarities among stem-associated tissues and subsequent divergence of root and leaf samples [45, 59].

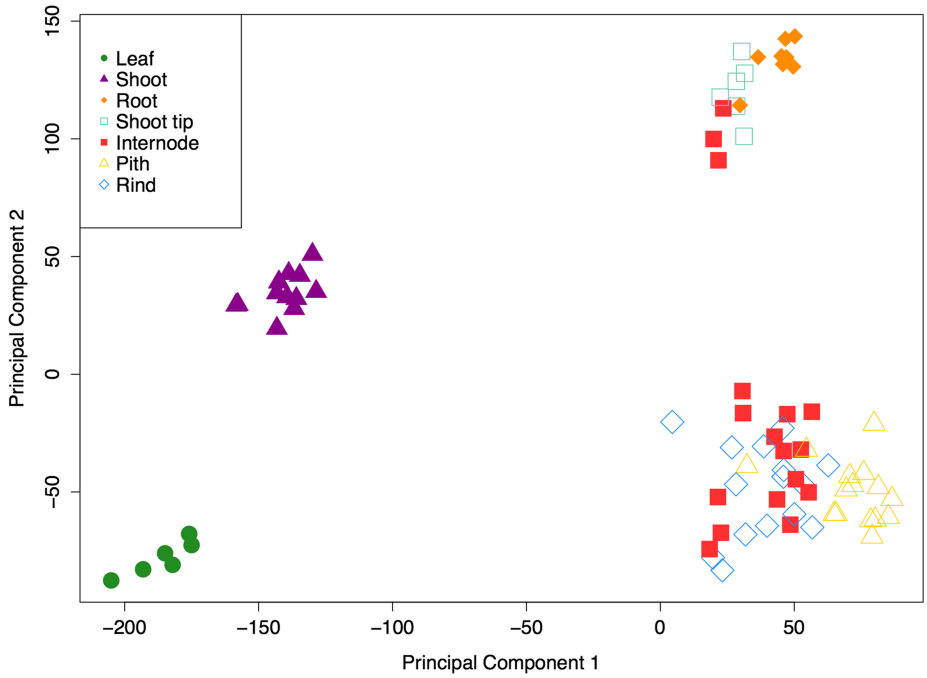


Figure 2.4 Classes sharing similar expression patterns. Principal component analysis was applied to 78 tissue samples, based on expression of 29,065 probe sets (27,577 genes, 654 controls and 834 small RNA probe sets). Each symbol represents a single sample. Tissue types are indicated by color and shape of symbol.

Interestingly, three out of 46 stem samples clustered near the group of meristematic tissues (roots and shoot tips). All three of these 'outlier' samples were collected at the top internode, 61 days after planting (DAP) in three of the six sampled genotypes (PI455230, PI152611, and AR2400). At 70DAP, the stem samples from same genotypes clustered with the other stem samples. These lines are characterized as high biomass genotypes, whereas the remaining three genotypes can be characterized as either grain or sweet lines (R159, Atlas, and Fremont). The PCA indicates that at 61DAP, the patterns of gene expression in the stem of the high biomass lines are more related to meristematic regions, or regions of active growth. While it is possible that these three stem samples were collected too close to the meristematic shoot tip region, further study may indicate that the differential transcriptome in the stems of these lines capture a transition zone of gene expression in which sorghum commits to post-reproductive pathways of sugar production and grain fill versus continued biomass production. This result further demonstrates the importance of examining genotype, tissue type, as well as temporal expression patterns when targeting transcriptional programs of interest.

Tissue and genotype-specific patterns of gene expression

To identify tissue-specific genes, we created genotype-specific datasets for PI152611, Fremont, and AR2400, each representing one of three major classes of sorghum: forage, sweet, and high biomass types respectively. Excluding replicate tissues from the same major organ, we identified genes exclusively expressed in the leaf, shoot, root, shoot tip and stem (Figure 2.5). The leaf and meristematic shoot tips expressed the greatest number of tissue-specific

transcripts across all three genotypes, whereas the seedling shoots expressed the fewest number of tissue-specific genes. Of particular interest in this dataset is the extent of variation observed across genotypes. For example, in stems, over 800 stem-specific genes are identified in representative examples of sweet and high biomass sorghum. Over 500 stem-specific genes are detected in forage sorghum; however, only 103 stem-specific genes are common among all three sorghum types. This lack of shared tissue-specific genes across genotypes is observed in all major tissue types.

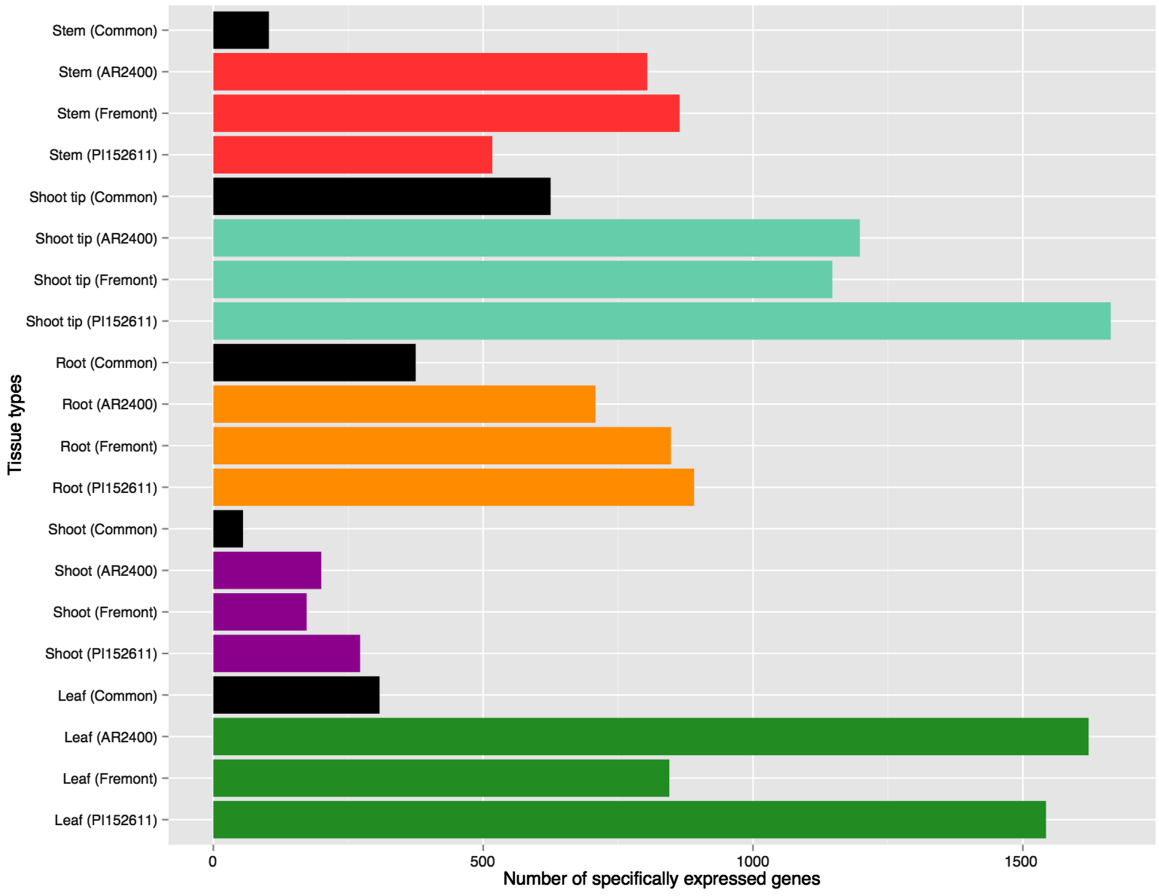


Figure 2.5 Number of tissue-specific genes in across sorghum ideotypes. AR2400: biomass sorghum; Fremont: sweet sorghum; PI152611: forage sorghum; Common: number of genes in common among all three ideotypes.

We also carried out this analysis for the small RNAs included on the array (Table A7). Similar to gene expression, we observed both tissue and genotype-specific expression of the small RNAs (Figure B5). For purposes of functional crop improvement, these result highlights the significance of intra-species variation in sorghum and the importance of selecting the appropriate genotype for targeted changes to gene expression via transgenic and breeding approaches.

To illustrate the expression dynamics among tissues, we also calculated the relative gene expression levels (Z-scores) of each of the major tissues (Figure 2.6). Consistent with previous studies, tissues with a relatively higher number of tissue-specific genes (e.g. leaf, root, shoot tip, pith) had a wide distribution of genes deviating from their mean expression. Stem-associated tissues had similar expression profiles and gene expression was closer to the overall average across tissue types [45, 59].

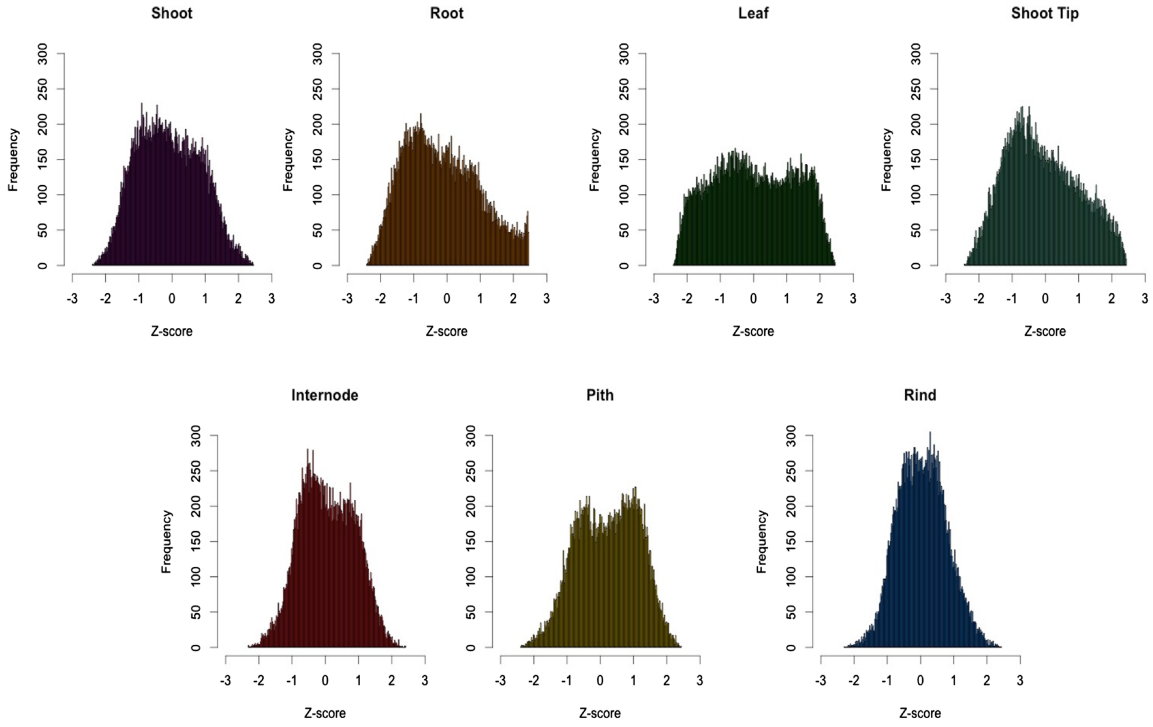


Figure 2.6 Distribution of global gene expression across sorghum tissue types. Histograms of relative expression levels (measured by Z-scores) in each tissue type. For each of these tissues, Z scores were calculated as follows: $Z = (X - X_{mean})/SD$, where X is the average expression of a given gene in a tissue, and X_{mean} and SD are the mean expression and standard deviation respectively of that gene across all the selected tissues.

We next attempted to determine whether functional gene classes were over-represented in specific genotypes. GO analysis did not reveal statistical differences in the enrichment of GO slim terms using agriGO (Fisher's exact test and the Yekutieli (false-discovery rate under dependency) multi-test adjustment method) [60]. However, this can partially be attributed to the incomplete annotation of the sorghum genome, as well as stage and tissue-specific expression not captured in our sample collection.

To identify genotype-specific expression patterns, we examined the expression of several known sugar metabolizing enzymes and sucrose transporters in sorghum with the hypothesis that differential expression of these genes would be observable across genotypes (Table A8). Differential expression between sweet and grain sorghum has recently been shown [42, 61], and our results further validate this observation, with the majority of sugar-related genes showing differential expression among tissues and genotypes. For example, sweet and high biomass varieties showed consistently higher expression of SPS2 and SPS5, sugar phosphate enzymes thought to play significant roles in sucrose biosynthesis, compared to grain varieties (Figure 2.7). A comprehensive gene list and more detailed expression analysis of sugar related genes across genotypes may provide insight into the mechanisms governing trade-offs in sorghum grain yield and stem sugar content.

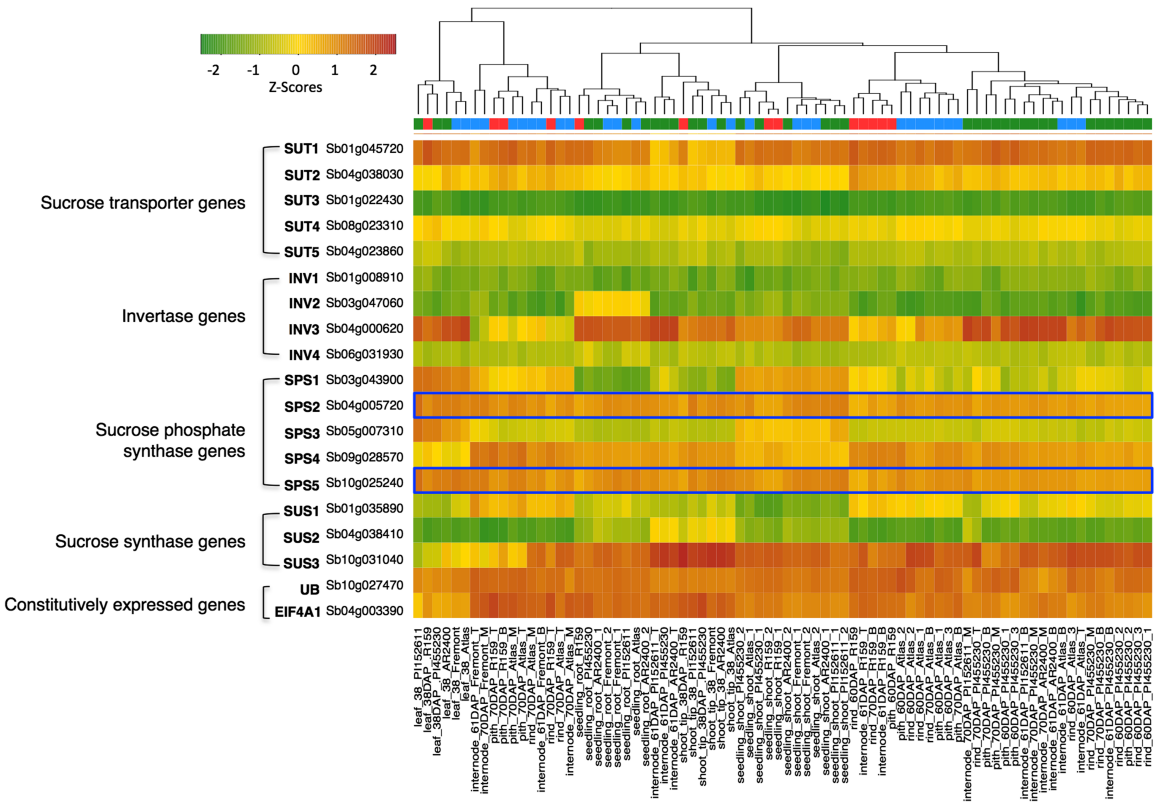


Figure 2.7 Hierarchical clustering of samples based on expression of sucrose metabolizing enzymes and sucrose transporter genes. Color bar key: Blue: sweet sorghum; red: grain sorghum; green: high biomass sorghum. Outlined in blue, the expression of sucrose phosphate synthase genes, SPS2 and SPS5, is consistently lower in grain types sweet and high biomass lines. Sugar metabolism gene list is appropriated from current literature [61].

We further analyzed tissue-specific transcripts to identify shared and specifically expressed genes in multiple tissues (Figures 2.8 and 2.9). To avoid variation in gene expression due to genotypic differences, we chose samples from the genotype Atlas for this analysis. We identified 587, 489, and 698 genes that are specifically expressed in leaf, stem and root and 232 and 688 unique genes that are expressed in shoot and shoot tips, respectively (Figure 2.8). We also identified 960 genes that are specifically expressed in stem rind (predominantly lignified sclerenchymatous cells) as compared to 928 genes that

are specifically expressed in stem pith (predominantly non-lignified parenchymatous cells; Figure 2.9).

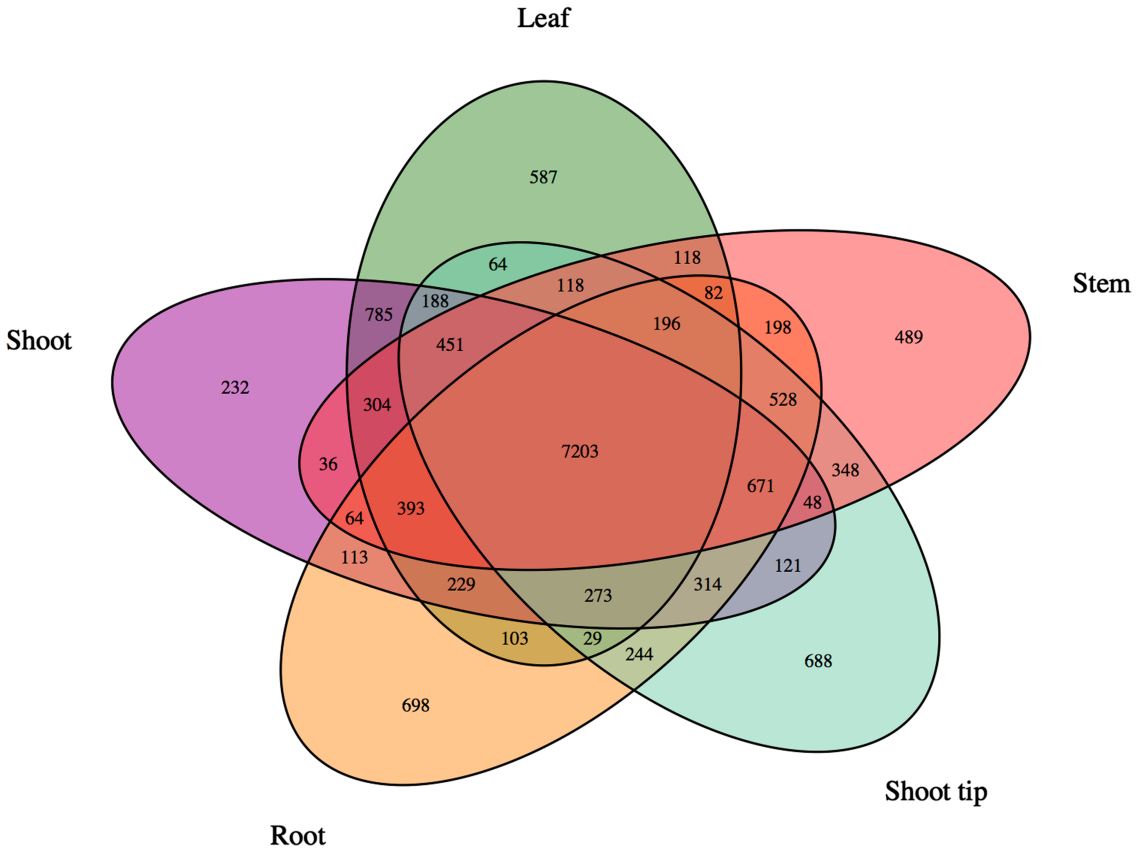


Figure 2.8 Number of shared and specific expression profiles of genes expressed in multiple tissue types (Atlas genotype).

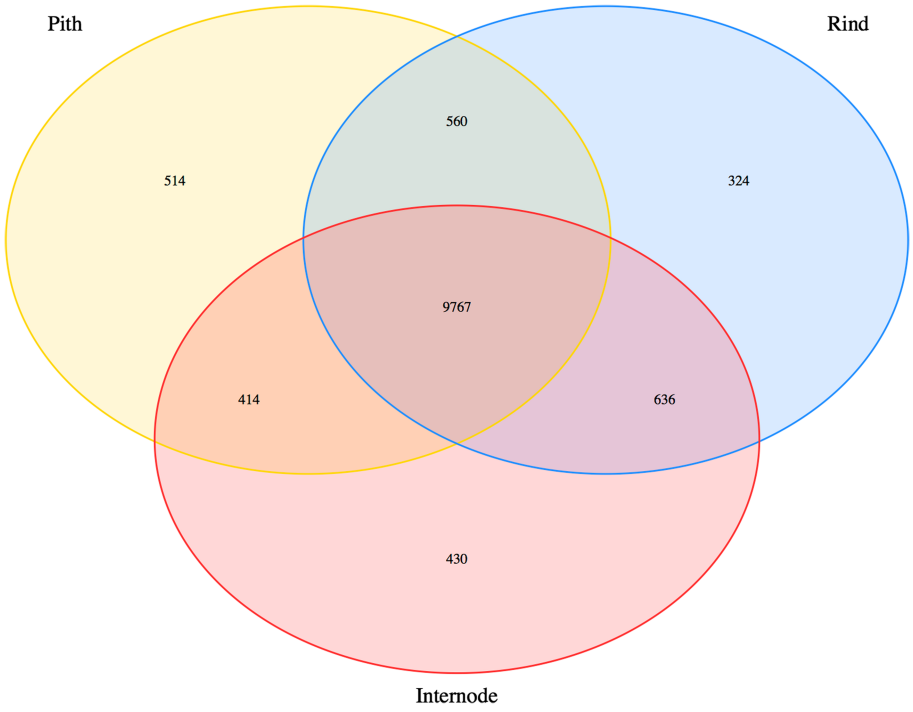


Figure 2.9 Number of shared and specific expression profiles of genes expressed in multiple stem tissues (Atlas genotype).

This dataset provides a unique opportunity to discover target sets of genes in core sorghum varieties that may be useful for modulating gene expression in a tissue-dependent manner. For example, these rind and pith-specific genes can be studied as potential candidate genes for biomass content and targets for compositional modification of biofuel feedstocks. Further, identification of promoter elements and corresponding DNA-binding regulatory proteins that regulate tissue-specific expression of genes could be identified from these data. As a direct application of this study, we are currently analyzing the promoter regions of candidate genes that are differentially expressed in the rind versus pith region of stem tissues.

Tissue-specific expression of genes involved in the phenylpropanoid-monolignol pathway

To exemplify the functional utility of this data, we highlighted the expression data of 10 key enzymes associated with the phenylpropanoid-monolignol biosynthesis pathway (Table A9). Currently, one of the primary strategies for bioenergy feedstock improvement is through lignin modification. Alterations in lignin content and composition aim to improve the digestibility of forage and saccharification efficiency of lignocellulosic biofuels [62, 63]. Thus, modifying the expression of genes in the lignin biosynthesis pathway is an attractive approach to achieving this goal.

Annotated in several databases, the majority of known and putative genes and homologs were analyzed for: phenylalanine ammonia-lyase (PAL, 9 sequences), coumaroyl shikimate 3'-hydroxylase (C3'H, 1), ferulate 5-hydroxylase (F5H, 3), cinnamate 4-hydroxylase (C4H, 3), 4-coumarate:CoA ligase (4CL, 5), cinnamoyl CoA reductases (CCR, 3), hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT, 1), caffeoyl-CoA 3-O-methyltransferase (CCoAMOT, 6), caffeic acid 3-O-methyltransferase (COMT, 1), and cinnamyl alcohol dehydrogenase (CAD, 1). Similar to previous studies in maize and switchgrass, the highest expression of these genes was found in the roots and stems [8, 64]. Further, hierarchical clustering reveals that the expression of lignin biosynthesis genes varies with developmental stage, as well tissue type and genotype (Figure 2.10). Distinct expression signatures of gene homologs as well as clustering of above-ground vegetative tissues according to developmental stage has precedence in maize and, in general, most of the lignin genes showed

organ-specific expression patterns consistent with studies in related species [45, 59].

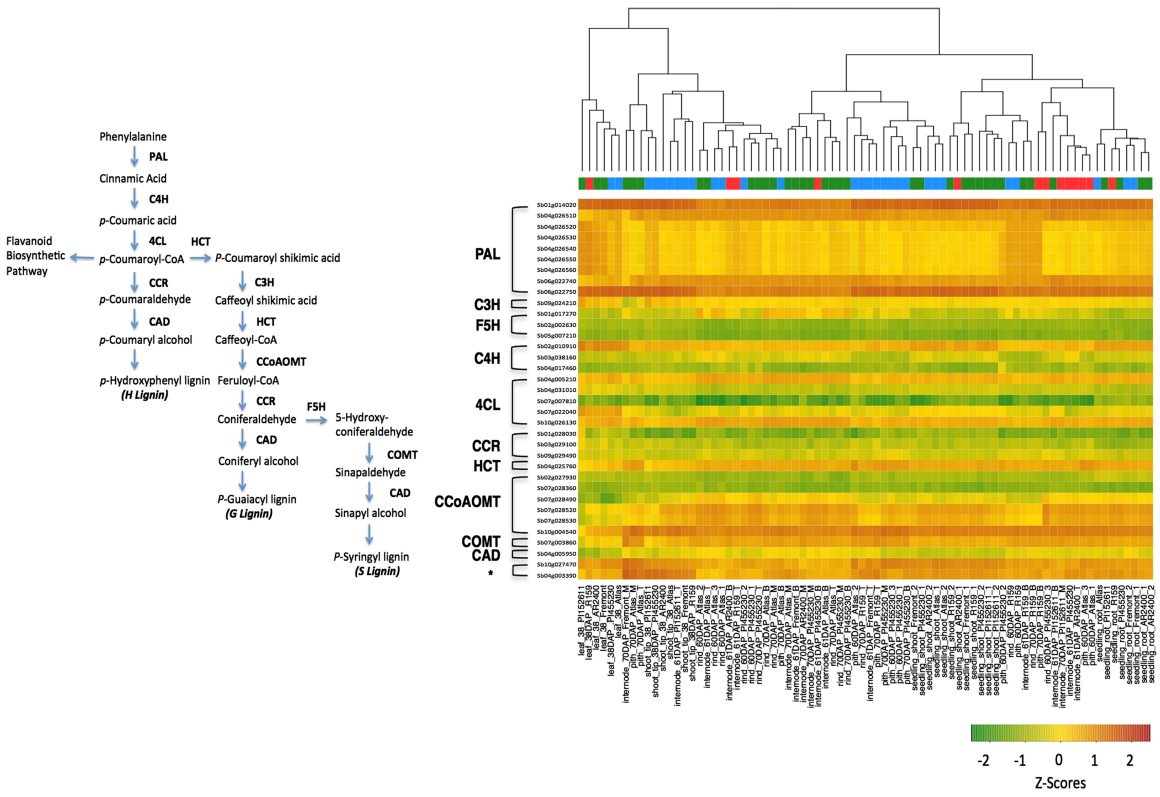


Figure 2.10: Hierarchical clustering of tissues based on expression of phenylpropanoid-monolignol biosynthesis pathway genes. *Constitutively expressed genes: Ubiquitin: Sb10g027470; EIF4A1: Sb04g003390. Color bar key: Blue: sweet sorghum; red: grain sorghum; green: high biomass sorghum. The color scale indicates the relative gene expression (Z-scores). Red, yellow, and green represent high, medium, and low levels of gene expression, respectively. The phenylpropanoid-monolignol pathway and enzyme nomenclature is appropriated from current literature.

2.4 CONCLUSIONS

Comprehensive transcriptome profiling provides a global overview of gene networks and allows for the discovery of functional connections between genes, mRNAs and their regulatory proteins. In the present study, we

constructed a gene expression atlas covering an array of tissues, developmental stages and genotypes using the first commercially available sorghum microarray (Sorgh-WTa520972F). We observed tissue and genotype-specific expression patterns of relevant metabolic pathways that highlight the significance of intra-species variation in sorghum.

Developed as a new resource for crop breeding and genomic discovery, Sorgh-WTa520972F is produced by Affymetrix and is available to the public research community.

2.5 METHODS

Tissue Collection

To study the sorghum transcriptome and build the present gene expression atlas, we collected 78 samples from various developmental stages and tissue types (Table A1). Six diverse sorghum genotypes were grown in Chromatin's greenhouse and field sites (Champaign, IL). These six genotypes were chosen to represent ideotypes of sorghum cultivation, including sweet, grain and high biomass sorghum varieties. Greenhouse grown seedling shoot and root samples were collected at 10DAP, which is roughly five days after plant emergence. Whole leaf and meristematic shoot tip samples were collected at 38DAP. This time-point captures the active growth phase of vegetative structures, including leaves, shoots and tillers. The stem tissue samples were collected at two time points: 61 and 70DAP. At 61DAP, the stem is fully formed in both flowering and non-flowering types. In flowering types, the head is also fully formed, and the period between 61 and 70DAP is a stage of active metabolism, capturing the transition between flowering (61DAP) and active grain filling (70DAP) [65]. The

stem tissue was further dissected into the pith and the rind. As a bioenergy crop, the majority of fermentable sugar available in sorghum is present in the pith. The majority of lignin, however, is found in the rind [66]. Two tissue types (shoot and root) were represented by two biological replicates.

Microarray Design

A whole-transcriptome exon array for *Sorghum bicolor*: *Sorgh-WTa520972F* was designed and utilized for the present expression study. The array contains 1,026,373 probes covering 149,182 exons (27,577 genes) across the *Sorghum bicolor* nuclear, chloroplast and mitochondrial genome. The sequences used to construct the probesets included all identified *Sorghum bicolor* exons from the Sbi1 assembly and Sbi1.4 annotation (http://phytozome.net). We also added sequences for putative non-coding RNAs in *Sorghum bicolor* that may play a role in gene regulation (*e.g.,* rRNAs, tRNAs, snoRNAs and microRNAs). Confirmed functional small RNAs in closely related species (corn, sugarcane) were also included in our array design (http://bioinformatics.cau.edu.cn/PMRD, http://ncrna.org/frnadb) (Table A7).

RNA Isolation and hybridization

Total RNA from all tissue types was extracted using a NucleoSpin RNA Plant Kit (Maxherey-Nagel, Germany). RNA integrity, as indicated by the detection of discrete ribosomal subunits, was verified electrophoretically. The RNA quality and quantity was further validated with a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). Prior to hybridization, the total RNA profile was also analyzed with Agilent 2100 Bioanalyzer (Agilent

technologies, Waldbronn, Germany). Synthesis of cDNA, probe labeling and hybridization was performed by Precision Biomarker (Precision Biomarker Resources, Inc. Evanston, Illinois)

Data extraction and evaluation of gene expression

Background correction and normalization were performed using a robust multi-chip average (RMA) algorithm in the Bioconductor *Affy* package [34]. Present calls for expressed genes were determined following established methods [45]. In brief, an expressed gene was identified by an RMA-normalized linear expression of >/= 320 in at least one of the 78 samples. The expression cut-off was five times the mean RMA-normalized signal from 576 negative-control oligos selected from the intronic regions of known constitutive genes (*e.g.*, actin, ubiquitin, and eIF4a1). A mean signal intensity of 64 was determined for the negative control oligos analyzed across all 78 slides. Constitutively expressed genes were identified by a RMA-normalized linear expression value of >/= 320 in all 78 samples.

Principal component analysis, hierarchical clustering and z-scores

To study the biological relatedness and identify expression trends among the samples, we utilized the cmdscale function and then plotted using R. We used RMA-normalized $\log_2$normalized expression values in the PCA analysis. Hierarchical clustering was performed using RMA-normalized $\log_2$normalized expression values and clustered using Pearson's correlation analysis. The Z scores were calculated as follows: $Z = (X-X_{mean})/SD$, where X is the average

expression of a given gene in a tissue, and X_{mean} and SD are the mean expression and standard deviation respectively of that gene across all the selected tissues.

GO Slim enrichment analysis

We evaluated enrichment of GO slim terms of biological process category (http://geneontology.org/GO.slims) in agriGO (http://bioinfo.cau.edu.cn/agriGO/) by Fisher's exact test (p-value ≤0.05) and the Yekutieli (false-discovery rate under dependency) multi-test adjustment method [60].

qRT-PCR

The relative mRNA expression was measured using Peltier Thermal Cycler PTC-200 PCR machine (MJ Research) and the SuperScript III Platinum SYBR Green One-Step qRT- PCR kit (Invitrogen). Three independent reverse transcription reactions were performed for each RNA sample, and qRT-PCR was carried out under the following conditions: 100 nanograms of each RNA sample was reverse transcribed at 60°C for 3 minutes, and reverse transcription was followed by initial activation at 95°C for 5 minutes, and 40 amplification cycles at 95°C for 15s and 50°C for 30s. Results were analyzed using MJ Opticon Monitor 3.1.32 software, and relative expression of mRNA was calculated by the comparative Ct method ($2^{-[\Delta][\Delta]Ct}$) [67]. Gene expression values across tissue types were normalized to ubiquitin expression.

Availability of supporting data

The transcriptome dataset supporting the results of this article is available through NCBI's Gene Expression Omnibus (GEO) under accession number GSE49879, and the Sorghum Genome Array is available through Affymetrix (http://affymetrix.com).

CHAPTER 3

INTEGRATING GENOME WIDE ASSOCIATION MAPPING AND IONOMICS TO IDENTIFY GENETIC LOCI AFFECTING SEED MINERAL CONCENTRATION IN *SORGHUM BICOLOR*

[2]

[2] Shakoor, Nadia, Ramesh Nair, Greg Ziegler, Brian Dilkes, Zachary Brenton, Rick Boyles, Erin Connolly, Stephen Kresovich and Ivan Baxter. " Ionomics and genome wide association reveal loci affecting seed mineral concentration in Sorghum bicolor." To be submitted to *Plant Physiology* or similar journal (December, 2014).

3.1 BACKGROUND / ABSTRACT

To address the challenge of global mineral malnutrition, current biofortification research in crop plants aims to improve mineral concentration and micronutrient bioavailablity via genetic and traditional breeding methods. Many staple food crops are also used as biofuels, and the chemical and mineral composition of these energy crops directly affect biomass quality and subsequent energy output. Identification of genes and QTL that impact mineral traits in the grains of major cereals, including sorghum, is fundamental to developing breeding and selection methods aimed at increasing bioavailable minerals and improving biofuel suitability and seed nutritional quality.

An ionomics approach, or simultaneous measurement of several key minerals, is well suited to identify the relationship between minerals, as well as to identify genes and QTL governing the accumulation of one or more minerals at the same time. We utilized the genotyped Sorghum Association Panel (SAP) and an additional 150 lines (CHP) representing elite, diverse bioenergy and sugar type sorghums to map QTL across 4 environments for seed mineral concentration. Genome-wide association analysis using a multilocus mixed model (MLMM) was applied to 20 macronutrients, micronutrients and elements of interest to plant and human health. Elemental analysis was carried out by ICP-MS.

Integration of ionomics and genome-wide association allowed us to identify several germplasm sources, genomic loci and candidate genes as potential biofortification targets for marker-assisted breeding and genomic selection. Specifically, we identify candidate genes implicated in the accumulation of manganese (Mn) and cadmium (Cd) in sorghum

seed. Manganese is an essential element necessary for plant and human nutrition, and with our ionomics approach, we were also able to identify significant phenotypic correlation between Mn and iron (Fe) content across multiple years and diverse environments. This suggests that common molecular mechanisms may underlie the uptake and metabolism of these important mineral traits. We also report high heritability for these elements that is encouraging to the development of breeding schemes for simultaneous improvement of these mineral traits in sorghum seed by marker-assisted selection.

Efforts to reduce toxic elements such as Cd in staple cereal crops via marker-assisted breeding are underway in rice [64], and here we provide analogous loci implicated in Cd accumulation in sorghum for functional characterization. In addition to candidates for these two elements, we provide a list of candidate loci in several other elements to be functionally characterized and discuss the application of powerful association mapping strategies towards compositional trait improvement.

3.2 INTRODUCTION

Sorghum is globally established as an important source of food, feed, sugar and fiber. Recent interest in bioenergy feedstocks also spotlights sorghum as an attractive alternative for sustainable biofuel production [4]. The mineral composition of stems, leaves and reproductive organs all significantly contribute to biomass quality. The seed bearing reproductive organs, or panicles, in sorghum represent up to 30% of the total dry matter yield [68]. Inorganic elements, particularly alkali metals, influence the combustion process and can

limit the effectiveness of conversion by contributing to slagging, fouling and corrosion of biomass furnaces and boilers [69, 70]. In order to achieve desirable end-use character of high biomass quality, agricultural strategies could potentially target reduction in specific minerals and compositional traits via transgenic and breeding approaches.

Modification of mineral composition and density in sorghum also has the potential to impact food biofortification strategies aimed at increasing the bioavailable minerals in the edible portions of the crop plant. Plant-based diets in which grains compose the major food source, require the availability of essential elements in the plant seed. Improved dietary mineral content in the seed can positively impact human and animal health. Fe and zinc (Zn) deficiencies affect over 2 billion people worldwide [71], and minor increases of these micronutrients in staple cereal grains, including sorghum, could potentially make a significant impact towards ameliorating the impact of these widespread deficiencies [72]. A lack of other essential minerals including magnesium (Mg), selenium (Se), Calcium (Ca) and Copper (Cu) are also prevalent in regions where grains, including sorghum and rice are cultivated and consumed as a staple crop [73]. While biofortification efforts strive to increase bioavailable essential minerals in the seed, similar strategies can also [74] be employed to reduce the concentration of toxic elements, including arsenic (As) and Cd.

The concentration of available minerals in the seed is dependent on complex metabolic, genetic and physiological factors that govern a suite of interconnected biological processes, including mineral uptake by the roots, translocation and remobilization within the plant, and ultimately import, deposition and assimilation in the seeds [75]. Spanning several large classes of

genes, the complex network governing seed mineral concentration in higher plants includes transporters, receptors, chelators, protein kinases, and transcription factors [75].

Previous studies have established that several minerals, including Fe, Mn, Zn, cobalt (Co) and Cd are coordinately regulated [76-78]. Further, ionomic signatures derived from multiple minerals have been shown to relate plant physiological status, including deficiencies in essential nutrients of Fe and P [79]. Examining the seed ionome as a whole has the potential to provide significant insight into the networks underlying ion homeostasis in sorghum.

With over 40,000 catalogued germplasm accessions, there is significant genetic variation and resources in sorghum that can be leveraged to identify genetic targets for modification of seed mineral composition [50]. While determining strategies to enhance or reduce mineral content for food or fuel, several components of seed mineral traits must be considered. These include: the heritability of the various mineral, genotype by environment interactions, and the availability of high-throughput mineral content screening tools [80] .

QTL mapping for seed mineral concentration has been carried out in a number of model species including Arabidopsis [75, 81], rice [82], wheat [83, 84] and maize [11]. There is considerable genetic and phenotypic diversity in sorghum, and genome-wide association (GWA) mapping is well suited for uncovering the genetic basic for complex traits, including seed mineral accumulation. One of the key strengths of association mapping is that *a priori* knowledge is not necessary to identify new loci associated with the trait of interest. Further, a GWA mapping population is comprised of accessions that have undergone numerous recombination events, allowing for a narrower

mapping interval. Previous GWA studies in maize [15], rice [16] and sorghum [25] have been successful in identifying the genetic basis for various agronomic traits. We utilized a similar mapping approach to identify potential loci underlying seed mineral accumulation in sorghum.

3.3 RESULTS

Representing sorghum diversity, we utilized two association panels of 407 and 150 genetically variant sorghum lines. The 407 accessions comprising the publicly available Sorghum Association Panel (SAP) were selected for genotypic diversity and phenotypic variation [85], and the 150 genotypes from Chromatin, Inc. (CHP) were selected based on agronomic traits and phenotypic diversity (Table A10). The SAP lines were previously genotyped using a genotyping by sequencing (GBS) method to identify new and informative SNP markers [25]. In order to obtain comparable genetic data for the complete mapping population, we also carried out GBS for the 150 CHP lines. The accessions in both individual panels are phenotypically, morphologically, and geographically variant and compose the genetic pool utilized in this study for dissecting the bases of mineral accumulation in sorghum.

Phenotypic diversity for seed mineral concentrations

We detected the effects of both genotype and environment on most of the elements measured (Figure 3.1), however the measured mineral concentrations of the present study corroborate the broad range observed in sorghum mineral literature [86-89].

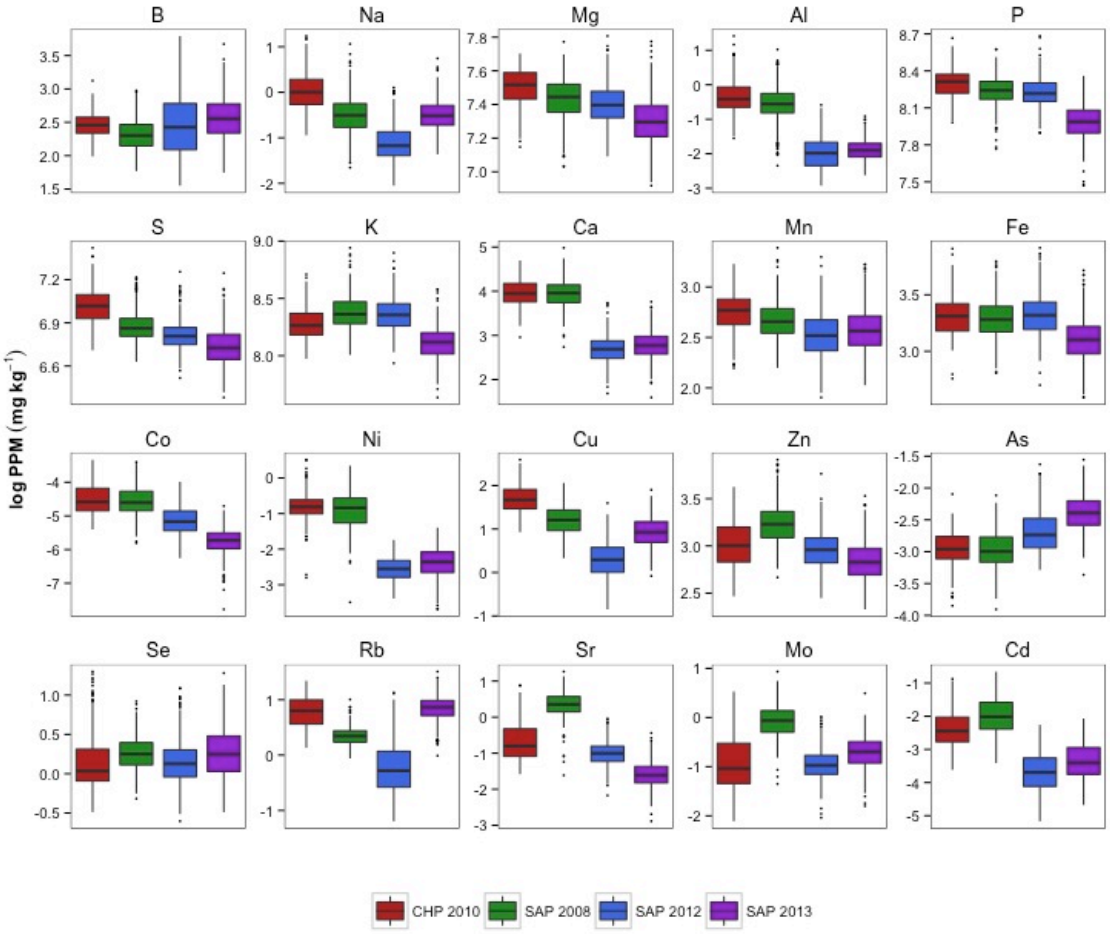


Figure 3.1: Box plots depicting the distribution of 20 minerals in four experimental populations. The raw concentration values for each of the minerals were log transformed to obtain normally distributed phenotypes.

Compared to micronutrients, the macronutrients (Ca, phosphorus (P), potassium (K), sulfur (S), sodium (Na) and Mg) in the study exhibited less phenotypic variation (Table A11). Similar to a study carried out in wild emmer wheat [90], grain Ca and Na showed the largest variation (nine and elevenfold, respectively) while the phenotypic variation in the remaining macronutrients ranged between 2.2 and 3.4 fold. With the exception of nickel (Ni) and aluminum (Al), seed micronutrient concentration showed phenotypic variation

ranging between 2.1 to 8.6 fold. High variation in these two elements may indicate a strong environmental effect on grain Ni and Al concentration or potential contamination during handling and analysis of the seeds [18].

We calculated broad-sense heritability to determine which proportion of the phenotypic variation in mineral concentration could be explained by the genetic variation present in the mapping panel (Table 3.1). Heritability estimates ranged from 23% (Sr) to 82% (Mn). We obtained high heritability (> 60%) for the majority of minerals including: boron (B), Na, Mg, P, S, K, Mn, Fe, Co, Cu, Zn, As, Se and molybdenum (Mo). Moderate heritability (30-60% was estimated for Al, Ca, Ni, (rubidium) Rb, and Cd. Low heritability (<30%) was estimated for strontium (Sr). The inheritance of seed mineral concentration is complex, and lower heritabilities could be explained by uncontrolled environmental difference between the replicated experiments. In mineral analysis of related species, low heritability has been reported previously for Al and Ni [82]. Low heritability for Na, Al, and Rb has also been shown in a similarly designed study in maize seed kernels [18].

Table 3.1 Mean, standard deviation, and broad sense heritability of seed mineral concentrations from the Sorghum Association Panel averaged across 4 environments. Mineral concentration values are presented as mg kg^{-1} and broad sense heritability (H^2) was calculated as described in the methods section. Data represents an average of individual samples (n=287) analyzed in 4 separate experiments. *Element concentration presented in $\mu g\ kg^{-1}$.

Trait	Sorghum Association Panel		H^2
	Mean	Standard Deviation	
B	*12.8*	*5.88*	*0.64*
Na	*0.572*	*0.316*	*0.61*
Mg	*1580*	*246*	*0.78*
Al	*0.277*	*0.269*	*0.38*
P	*3350*	*623*	*0.59*
S	*890*	*127*	*0.75*
K	*3850*	*795*	*0.61*
Ca	*25.3*	*18.8*	*0.33*
Mn	*13.5*	*3.33*	*0.82*
Fe	*24.9*	*5.7*	*0.76*
Co	*6.11**	*4.48**	*0.51*
Ni	*0.18*	*0.187*	*0.36*
Cu	*2.54*	*1.21*	*0.70*
Zn	*19.9*	*5.71*	*0.66*
As	*0.0796*	*0.0361*	*0.61*
Se	*1.32*	*0.454*	*0.80*
Rb	*1.76*	*0.848*	*0.37*
Sr	*0.573*	*0.586*	*0.23*
Mo	*0.603*	*0.306*	*0.53*
Cd	*0.0683*	*0.0739*	*0.42*

We used two different approaches to identify elements under shared control, correlation analysis and principal component analysis (PCA) (Figure 3.2, Figure 3.3 and Table A12). Highly correlated minerals across both association panels included Mg-P, Mg-Mn, P-S and Mg-S. Divalent Ca^{2+} and Sr^{2+} are chemical analogs and strong correlation was observed as has previously been reported in other species [91-94]. In both the SAP and CHP panels, the first two principal components accounted for a large fraction of the variance (36% and 37%

respectively). Distinct clustering of known elemental relationships, including Ca and Sr can be observed. A cluster of the essential metal micronutrients, Fe, Zn and Cu is distinguishable, and clustering of Mg and P is consistent with previous studies in wheat that show core similarities among families of known mineral genes. Seed P is predominately stored as inositol-hexaphosphate IP_6 or phytic acid. Phytic acid is strong chelator of cations, including Mg^{2+}, and may explain the significant positive relationship identified in this study [95, 96].

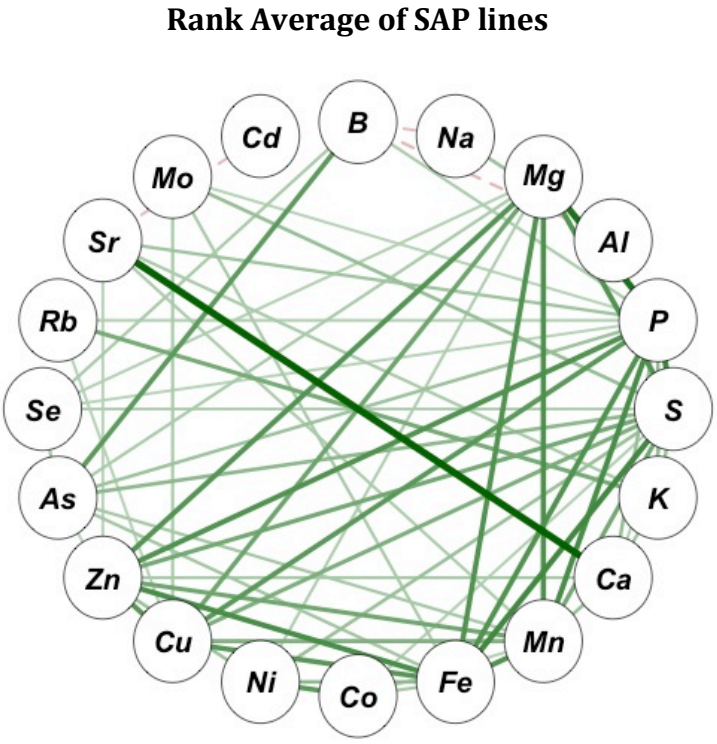


Figure 3.2 Correlation network of seed mineral concentrations using rank average data calculated across replicates from SAP association panels. Green solid lines represent positive correlation values. Red dashed lines represent negative correlation values. Intensity and thickness of lines indicate degree of correlation. Mineral correlation values can be found in Table A12. Correlation networks for SAP 2008, SAP 2012, SAP 2013 and CHP 2010 can be found in Figure B7.

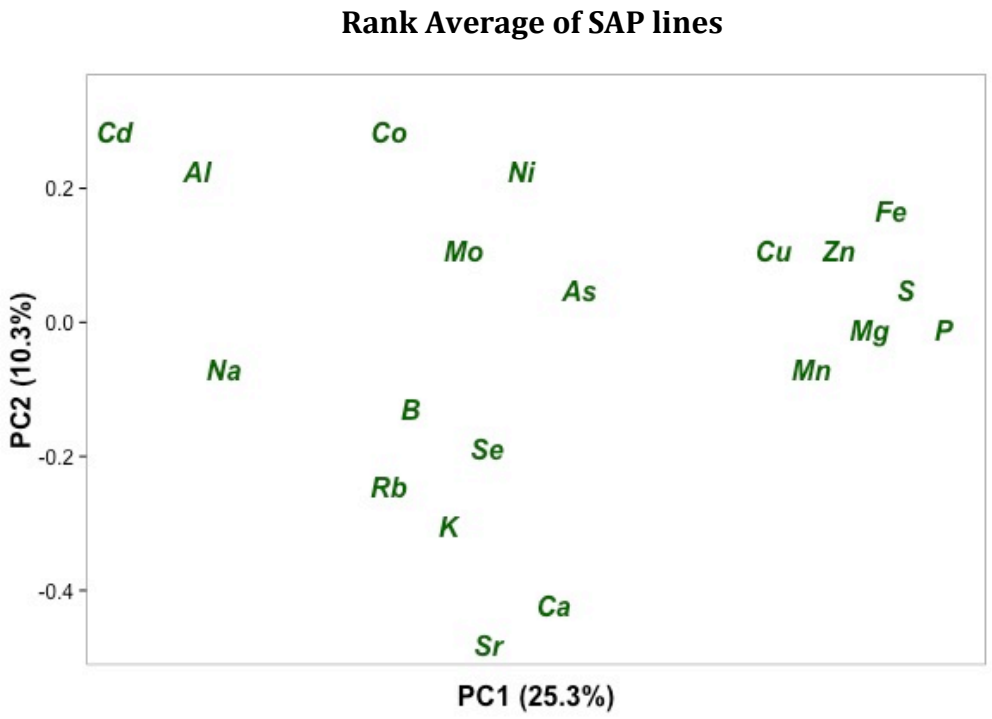


Figure 3.3 Principal component analysis applied to the rank average seed concentrations for 20 minerals in the SAP lines across experiments. Each symbol represents a single element. PCA analysis for SAP 2008, SAP 2012, SAP 2013 and CHP 2010 can be found in Figure B8.

Genome-wide association mapping of seed mineral traits

To dissect the genetic basis of natural variation for seed mineral concentration in sorghum seed, GWA mapping was performed independently on two diversity panels, the Sorghum Association Panel and the Chromatin 150 panel using both an optimal model obtained from the multi-locus mixed model (MLMM) algorithm and a compressed mix linear model (CMLM) to account for population structure. With P values below a Bonferroni-corrected threshold (P = 0.05), we identified a large number of SNPs significantly associated with seed mineral concentration using both approaches. Significant SNPs identified with the MLMM approach were prioritized for further analysis (Tables A13, A14, A15

A16 and A17). Compared to traditional single-locus approaches (e.g. CMLM), MLMM utilizes multiple loci in the model, which contribute to a higher detection power and lower potential of false discoveries [23]. Briefly, MLMM is based on EMMA [24] and relies on the iterative use of a simple K, or Q+K, mixed-model algorithm. At each step of the MLMM, the phenotypic variance is divided into genetic, random and explained variance. The most significant marker is included as a cofactor, and the variance components of the model are recalculated. With each successive iteration, the genetic variance approaches zero, and an optimal model including cofactors that explain the genetic fraction of the phenotypic variance is determined.

In an effort to comprehensively identify significant SNPs associated with mineral concentration, we created several datasets for GWA analysis. GWA scans across individual experiments identified 270, 228, 207, and 355 significant SNPs for all twenty mineral traits in the SAP2008, SAP2012, SAP2013 and CHP2010 panels, respectively (Tables A13, A14, A15, and A17). For the SAP experiments, we also carried out a rank transformation method by ranking the individual lines of each experiment by mineral concentration and carrying out GWAS on the average rank across the four SAP environments [97]. We identified a total of 255 significant loci in the ranked dataset for the twenty mineral traits (Table A16).

To identify common SNPs across the multiple SAP locations, we identified shared significant SNPs from each environment that were present in 2 or more environments (Table A18). For example, GWA for Ca concentration in all three of our SAP experiments identified significant SNPs within 5kb of locus Sb01g008440 on chromosome 1. Sb01g013660 is a putative calcium homeostasis

regulator (CHoR1) [98]. We also identified several significant SNPs that colocalized for multiple mineral traits (Table A19).

The GWA results for each mineral trait in all four datasets were individually examined. However, since the goal of our study was to broadly capture markers for each of the mineral traits in multiple environments, we focused on the GWA results from the SAP rank average dataset and compiled the GWA results for each mineral trait obtained at the optimal step of the MLMM model (Figure B9). Here we highlight the results for Cd; GWA across multiple environments identified one significant SNP (S3_8883293) associated with Cd levels. (Figure 3.4). Nominal inflation in the distribution of expected vs. observed P values suggests that population structure was well-controlled and false positive association signals were minimized (Figure 3.5). At each step of the MLMM analysis, the phenotypic variance is divided into random, genetic and explained variance. The stopping point for including cofactors occurs as the genetic variance approaches zero. The optimal MLMM model (ExtBIC) included one SNP on chromosome 2 that explains 18% of the phenotypic variation in cadmium (Figure 3.6), and the allelic effects of each genotype were estimated (Figure 3.7). This significant SNP on chromosome 2 (S3_8883293) is in LD with a homolog of a well-characterized cadmium transporter in plants, heavy metal ATPase 2 (HMA2).

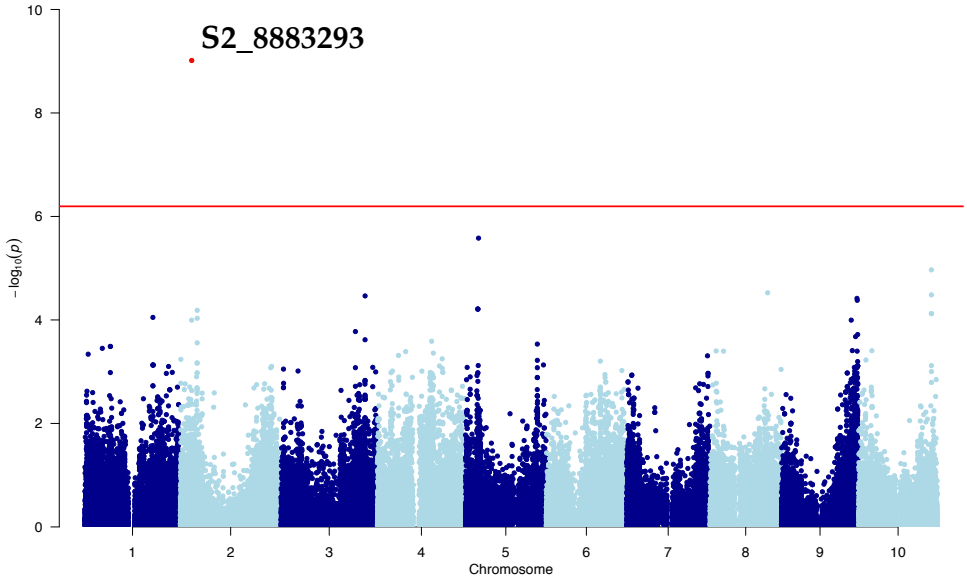


Figure 3.4 Manhattan plot displaying Cd GWAS results ($-\log_{10}$(P)) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis). The red lines indicates a Bonferroni-corrected threshold of 0.05.

.

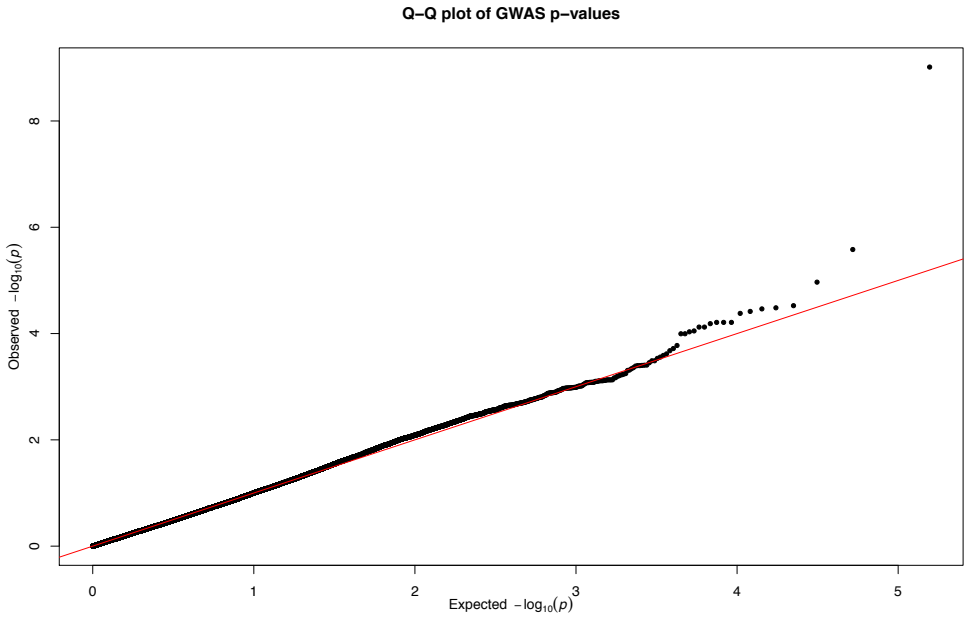


Figure 3.5 Quantile-quantile (Q–Q) of observed P-values against the expected P-values from the GWAS analysis for mineral concentration

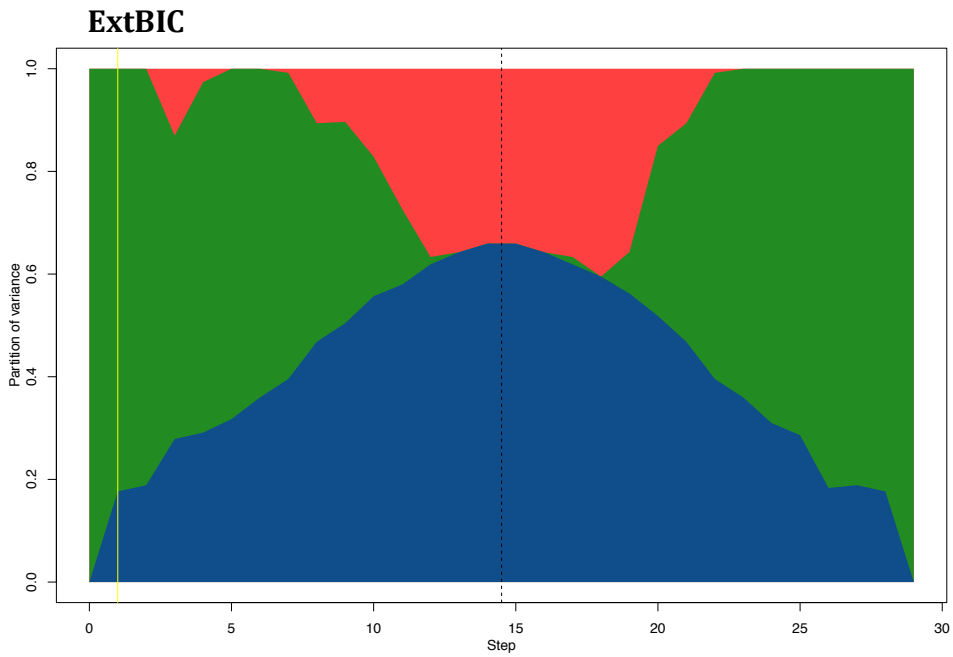


Figure 3.6 Evolution of genetic variance at each step of the MLMM (blue, genetic variance explained; green, total genetic variance; red, error) for the optimal model (step indicates extended Bayesian information criterion (ExtBIC))

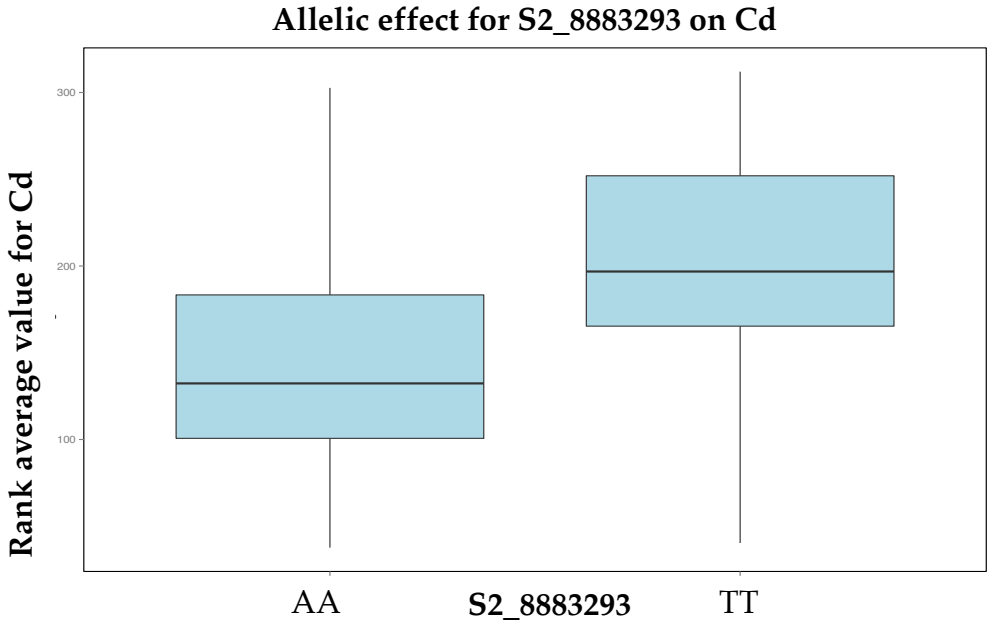


Figure 3.7 Allelic effect for the significant SNP marker on chromosome 2

3.4 DISCUSSION

Increasing the concentration of minerals essential for human and animal nutrition (e.g. Fe and Zn) while simultaneously minimizing anti-nutrients and toxic elements (e.g. As and Cd) is a significant goal of the global agricultural community. In the last decade, molecular genetic and traditional breeding methods in agriculture have been systematically applied towards the development of micronutrient-enriched staple food crops [80] Mineral homeostasis in plants, however, is extremely complex and both genetic and environmental factors can reduce the bioavailability of nutrients [99] . Ionomic analysis, or concurrent measurement of multiple elements, allows for higher resolution of the regulatory network governing mineral concentration. Through the quantitative and simultaneous measurement of an organism's elemental composition, we are provided with a snapshot into the functional state of an organism under different experimental conditions [100].

Genotype, environment, soil properties, and nutrient interactions can all significantly impact the mineral concentration in sorghum grain. However, in this study, we observed high heritability of several elements and report general homeostasis of mineral concentration across very diverse environments (Figure 3.1 and Table 3.1). High heritability of a trait of interest is desirable when developing breeding strategies for crop improvement, and if selection strategies include improvement of multiple traits, phenotypic correlations between traits are also relevant. The heritability calculations and the relationships we report between important mineral elements are encouraging for the development of breeding schema for improved mineral profiles in sorghum. In the present

study, observed correlations of several elements may indicate that an increase or reduction in one or more minerals may simultaneously affect the concentration of other minerals present in the seed (Figure 3.2). Further, these correlations suggest the potential for uncovering common genetic associations for multiple elements as we identified colocalized peaks for several mineral traits (Table A19). For example, several significant SNPs colocalized for strongly correlated Ca and Sr (r= 0.79) as well as Mg and P (r= 0.71). Shared SNPs and colocalization of significant loci across multiple mineral traits suggests the possibility of tightly-linked genes or genes with pleiotropic effects and has been documented in recent GWA studies, including experiments in tomato [101] and rice [102].

Candidate genes

One of the primary goals of this study was to utilize GWA analyses to identify candidate genes and novel loci implicated in the genetic regulation of sorghum seed mineral traits. We identified numerous significant SNPs for all twenty mineral traits that currently do not associate with functionally annotated genes. Although it is likely some fraction of these are false positives, many may be novel associations with as-yet undiscovered causal genes and merit further investigation. We did, however, identify several significant SNPs that fall directly within a characterized candidate gene or are in close proximity, or LD, with putative candidates.

Manganese

One of the essential minerals critical to human and animal nutritional requirements is Mn. Associated with amino acid, lipid and carbohydrate

metabolism, manganese is necessary for suite of processes including immune function, reproduction, digestion and bone development [103].

We identified significant GWAS associations in the putative sorghum homolog for MTP11 (Sb03g039220). MTP11 is a protein in the cation diffusion facilitator (CDF) family that function in the transport of metal ions. AtMTP11 expressed in *Saccharomyces cerevisiae*, conferred Mn tolerance and was reported to transports Mn^{2+} via a proton-antiport mechanism [104].

Cadmium

The seeds of staple crops are a major source of essential nutrients, however, they also can be a source of toxic heavy metals, including cadmium. Itai-itai disease is associated with cadmium toxicity due to dietary consumption of rice with high seed concentration of Cd [105]. Contamination of ground water and subsequent uptake and absorption by the plant yield high levels of Cd contamination in the seed [106]. GWA analysis identified significant SNPs associated with a paralogous set of cation-transporting ATPases (Figure 3.3), Sb02g006940 and Sb02g006950. Both of these sorghum genes are orthologs of Arabidopsis HMA genes in the heavy metal–transporting subfamily of the P-type ATPases. Respectively, Sb02g006940 and Sb02g006950 are 51% and 52% identical to Arabidopsis HMA3 and 40% and 42% identical to HMA2. HMA2 and HMA3 in Arabidopsis are tandem genes on chromosome 4, and a recent study in Arabidopsis revealed that HMA3 is the major loci controlling natural variation in leaf cadmium [107]. These sorghum genes are strong candidates for HMA2-3, and merit functional characterization.

Nickel

In addition to Ca, Cu, Fe, K, Na, Mn, Mg, Cu and Zn, Ni is an essential nutrient required for plant growth. However, similar to Cd, high Ni concentrations in soil can be toxic to the plant, resulting in reduced biomass and crop yield. Also, similar to other heavy metals, including Cd, Co Cr, Pb and Zn, Ni can interfere with the uptake of Fe, an important essential nutrient for the plant and a desirable mineral for seed accumulation [108] .

The most significant SNP for Ni concentration in the SAP 2008 environment was S6_53175238. Also in the top 20 SNPs for the SAP 2012 environment and a significant hit in the rank average analysis, this SNP is in LD (<15kb) with the gene Sb06g024040. Annotated in Phytozome (http://www.phytozome.net/) as "similar to Iron transport protein 2", Sb06g024040 is a sorghum homolog of the Yellow Stripe-Like 3 (YSL) family of proteins. Originally identified in maize, the YSL proteins are a subfamily of oligopeptide transporters involved in metal uptake, homeostasis and long-distance transport [109]. YSL3 is suggested to transport metals bound to nicotianamine (NA)[110], and in *Thlaspi caerulescens* (a metal hyper-accumulator plant), YSL3 functions as Ni–NA influx transporter [111].

Molybdenum

A vital component of the molybdoptrin cofactor, Mo is an essential trace element in plants and animals necessary for enzymatic processes, including nitrate assimilation, sulfite degradation and abscisic acid biosynthesis [112]. Mo concentration in the seed has also been implicated in improving seedling vigor [113, 114]. GWA mapping across experiments in the SAP identify multiple significant SNPs markers directly within or in linkage disequilibrium (LD) with

locus Sb03g036740 on chromosome 3. This protein has not been characterized in sorghum, however GO annotation and ortholog analysis suggest homology to proteins classified under the sulfite exporter TauE/SafE family of proteins. Sulfur assimilation for cysteine biosynthesis in plants occurs via sulfite conversion to sulfide. Sulfite oxidase (SO) belongs to the class of molybdenum cofactor- containing enzymes, and plays a role in the oxidation of excess sulfite that is produced when sulfur-containing structures are broken down, or when plants are exposed to SO_2. In order to maintain metabolic homeostasis, an interactive relationship between SO and sulfite exporters can be postulated.

3.5 CONCLUSIONS

In the present study, we utilized GWA mapping to identify a number of genetic loci and candidate gene associations for future study. We identify co-localization of significant SNPs for different elements indicating the potential association of specific genomic regions with related physiological processes in metal uptake, transport, traffic and sequestering. The use of a multi-element, or ionomic approach to the analysis allows for the identification of SNPs that confer multiple advantageous traits that can be selected for in breeding programs. Fine mapping using targeted resequencing approaches can also be utilized to identify and confirm the causal polymorphism underlying specific mineral traits.

Additional studies of the SAP and CHP mapping populations in multiple environments would be informational in determining the contribution of environment to the total variation of the individual mineral trait. Integrating ionomic and genomic analyses to identify and and understand the molecular mechanisms underlying mineral profiles is imperative to developing techniques

for targeting genes and modifying gene expression in the mineral accumulation pathways via transgenic and breeding approaches.

3.6 MATERIAL AND METHODS

Plant material

Development of the Sorghum Association Panel has been previously described [85]. Seeds from 407 lines that comprise the Sorghum Association Panel (SAP) and 150 lines from the 'Chromatin 150 Panel' (CHP) were utilized for this study. The SAP 2008 and SAP 2009 seeds were obtained from Germplasm Resources Information Network (GRIN) and were produced in Lubbock, Texas (USDA-ARS Cropping Systems Research Laboratory) in 2008 and 2009. The SAP 2012 seeds were produced in Puerto Vallarta, Mexico in 2012. The SAP 2013 seeds were produced in Florence, SC in 2013. The CHP seeds were produced in Puerto Rico in 2010 and were provided by Chromatin, Inc. Mineral concentration in a total of 4 seeds was determined by ICP-MS after nitric acid digestion.

Phenotypic Elemental Analysis

Four seeds per accession were weighed from each panel and a minimum of two replicates from each accession of the SAP 2008, SAP 2013 and CHP 2010 panels were analyzed by ICP-MS. All samples were digested with 2.5 mL of concentrated nitric acid. Elemental analysis was performed with an ICP-MS for B, Na, Mg, Al, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd following established protocols [115]. A reference was derived from a pool of sorghum seed samples and was run after every 9th sample to correct for ICP-MS run-to-run variation and within-run drift.

Genotyping by Sequencing

Genomic DNA from CHP lines was extracted from seedlings approximately 7 days after germination using a modified CTAB protocol and quantified using PicoGreen (Invitrogen, CA). Isolated DNA was submitted to Cornell University Institute for Genomic Diversity for execution of the GBS pipeline [116]. Briefly, reduced-representation DNA libraries were constructed using ApeK1 restriction enzyme, sequencing was carried out in 96-plexes on the Illumina HiSeq2000, and the data was processed using the TASSEL GBS pipeline. GBS markers for the SAP lines used in this study have been previously described [25].

Data Processing and Analysis

Phenotype data were generated for 407 SAP and 150 CHP lines. After removing SNPs with more than 20% missing data and minor allele frequencies below 0.05, genotype data for these lines includes 78,012 SNPs and 41,464 SNPs for the SAP and CHP panels, respectively. Broad sense heritability was calculated from two-way analysis of variance from the estimates of genetic (σ^2_G) and residual variance (σ^2_e) derived from the expected mean squares as $H^2 = (\sigma^2_G)/ (\sigma^2_G - \sigma^2_e)/k$ where k is the number of replications. To ensure normality in the data distribution of the phenotype, the Box-Cox procedure was carried out on the phenotype scores to identify the best transformation method [117]. The 'boxcox' function in the MASS package in R was utilized to carry out the appropriate transformations [118, 119]. In order to address potential confounding factors in the GWA analysis, specifically ICP run-to-run variation and the weight correction calculation, we used linear regression to compute residuals adjusted for ICP run and sample

weight. These residuals were used to test for association with qualifying SNPs in the GWA analysis.

GWAS

GWAS was executed in R using Genomic Association and Prediction Integrated Tool (GAPIT) using GLM as well as CMLM to account for population structure [120]. Significant associations were determined by estimates of false discovery rate (FDR) (P = 0.05) [121]. The MLMM procedure was subsequently implemented to resolve potential false positives and increase the detection power of association signals [23]. We utilized MLMM to perform a two-model selection procedure, the extended Bayesian information criterion and the multiple-Bonferroni criterion, to identify the optimal model. We utilized a threshold of genome-wide significance threshold α= 0.05 after Bonferroni correction.

Acknowledgments

This project was partially funded by the iHUB Visiting Scientist Program (http://www.ionomicshub.org).

Availability of supporting data

The datasets supporting the results of this article are available through Purdue Ionomics Information Management System (PiiMS) at http://www.ionomicshub.org

CHAPTER 4

CONCLUSION

4.1 CONCLUDING REMARKS AND AREAS FOR IMPROVEMENT

The data generated and analyzed for the completion of this thesis was intended to provide novel insight into the genetic architecture underlying the numerous and diverse traits that characterize sorghum as a valuable crop for global fuel and feed. With every major crop, a large-scale transcriptome atlas is available to the research community for the examination of gene expression patterns on a genome scale. Such a resource was not previously available to the sorghum research community, so we custom-designed a whole-genome gene expression arrays in collaboration with Affymetrix that targeted 149,182 exons (27,577 genes) across the Sorghum bicolor nuclear, chloroplast, and mitochondrial genomes. We built an extensive gene atlas using a panel of 78 RNA samples derived from a combination of six sorghum genotypes representing grain, sweet, forage, and high biomass ideotypes (R159, Atlas, Fremont, PI152611, AR2400 and PI455230), four different tissue types (shoot, root, leaf and stem), and two dissected stem tissues (pith and rind). In Chapter 2, we present a summary of the microarray dataset, including analysis of tissue-specific gene expression profiles and associated expression profiles of relevant metabolic pathways.

With an aim to enable identification and functional characterization of genes in sorghum, this expression atlas presents a new and valuable resource to

the research community. Further research, with the inclusion of additional samples would increase the utility of this resource. With an original aim to identify tissue-specific promoters in the stem, we focused exclusively on the vegetative structures. A more complete gene atlas could be generated with the inclusion and analysis of sorghum reproductive tissues. Also, for the sake of time and cost, we utilized only one or two samples of each tissue and genotype. Biological and technical replicates of all the experimental samples would confer more robust results and higher confidence in the conclusions of the analysis [122].

This microarray can be utilized to identify differential gene expression related to key metabolic processes (*e.g.*, starch/lignin biosynthesis) for the identification of regulatory regions. Additional avenues for future study with this array are wide-ranging and can include gene expression profiling during abiotic/biotic stress, plant infection and disease establishment to investigate genetic mechanisms and applications to plant breeding and crop improvement. Detailed expression analysis of small RNAs included in the array design may also reveal key insights in diverse biological processes, including RNA-guided gene regulation. Sorgh-WTa520972F can also be utilized in quantitative trait locus (QTL) mapping and validation methods (*e.g.*, identify differentially expressed genes from 'tolerant' versus 'sensitive' varieties). Minimal costs associated with microarray analysis allow for the generation of high-throughput expression profiles or combinations of profiles of elite breeding lines for accelerated crop-breeding efforts. Applications of this resource can target numerous agronomic traits in sorghum as well as provide insight in closely

related grasses (*e.g.*, sugarcane, switchgrass, *Miscanthus x giganteus*) for improved feedstock development.

Presented in Chapter 3, we generated and analyzed the seed ionomic profiles and genetic data for a large panel of diverse sorghum lines, including 400 genetically diverse lines (SAP) and an additional 150 lines (CHP) representing elite, diverse bioenergy and sugar type sorghums. We examined the seed ionome of each of the lines from the SAP across multiple locations to gain insight into the networks underlying ion homeostasis in sorghum. While our study utilized multiple locations, we had to reconcile the confounding factor of multiple years and a lack of biological replicates. We also did not analyze soil samples in each of these locations. Field sites, including the ones in Florence, SC are tested for soil fertility ever year and this data could be incorporated in our analysis for these locations. Variability in the soil mineral content across our diverse locations may have affected the mineral concentration present in our seed analysis. Tissues other than seeds can provide information on nutrient status, and ionomic sampling of leaves or roots, might provide additional insight. Validation of candidate SNPs can also be carried out by raising the SAP genotypes under a set of common environments ('common garden' experiments). This would allow for a more accurate calculation of heritability and provide strong evidence that mineral concentration is a highly heritable trait in which the majority of the phenotypic variation in the population can be attributed to the additive effects of allelic variation at mineral-related loci among the experimental population.

4.2 FUTURE DIRECTIONS

We can integrate the genome-wide association mapping techniques developed in Chapter 3 to identify loci affecting additional compositional traits in *Sorghum bicolor*. Cell wall components of forage grasses provide a high amount of convertible energy for ruminant animals. Grass cell walls in particular, are comprised of a matrix of hemicellulose, glucose and lignin. The total available energy is dependent on the digestibility of these cell wall components, and the lignin and phenolic acid fraction of the cell wall matrix negatively impact digestibility [26]. Accelerated breeding for high digestibility via reduced fiber and lignin concentration is a primary goal for forage crop improvement. Modifying forage digestibility necessitates the identification and understanding of the components of the lignin biosynthetic pathway.

In an effort to address these requirements for forage sorghum compositional trait improvement, we have briefly outlined a future study for which preliminary data has been collected.

Study Title

Integrating genome-wide association mapping and NIR analysis to identify loci affecting stalk compositional traits in *Sorghum bicolor*

Collaborators/Co-authors: Stephen Kresovich, Todd Mockler, Scott Staggenborg, Owen McSpadden, Rick Boyles, and Zach Brenton

Aims

We utilized the same SAP and CHP lines from this thesis that represent elite, diverse bioenergy and sugar type sorghums to map QTL for traits relevant to cellulosic biofuel production, including ash content, neutral detergent fiber digestibility (NDFD), glucan, xylan, and lignin concentration. Similar to the mineral GWAS study of Chapter 3, the specific objectives of this goal will be to 1) locate QTL for biomass traits with GWA analysis and 2) identify genotypic correlations, between-trait correlations and QTL colocation.

Experimental methodology

407 lines from the 'Sorghum Association Panel' (SAP) and 150 lines from the 'Chromatin 150 Panel' (CHP) were planted on May 2012 and May 2013 at the Pee Dee Research & Education Center (Florence, SC). The precise location of the field at the station shifted in 2013, however, the soil type at both locations was sandy loam- a soil type consisting of less than 7 percent clay, less than 50 percent silt, and between 43 and 52 percent sand [123]. Plots in 2012 and 2013 were arranged in a randomized complete block design with two replications per year. In 2012, plots were harvested at physiological maturity, and a representative sample of 1kg of aboveground biomass was harvested per plot for NIR analysis.

In 2013, we harvested two replicates of the SAP, CHP, and an additional 400 lines were included in 2013 for bioenergy characteristics (e.g. biomass, bmr, etc.) for a total of ~1500 samples. In addition, we also collected samples from the SAP and CHP (~1100 samples) at 60 days after planting. The rationale for the 60 day sampling is that would like to uncover correlative traits across vegetative

and harvest samples. Whole plant samples were dried at 60°C for approximately 5 days, and dried samples were ground with a Wiley mill to a 2-mm particle size for NIR analysis. Samples were scanned through a near-infrared reflectance spectrophotometer (NIRS). FT-NIR spectra were recorded on a multi-purpose analyzer (MPA) spectrometer (Bruker Optics, Wissembourg, France). The MPA is software-controlled by the OPUS software Version 7.0 which is provided by Bruker Optics.

Further, in collaboration with Dr. Stephen Kresovich (Clemson University) and Dr. Andrew Paterson (University of Georgia), we have analyzed three F3 NAM populations via NIRS. There are approximately 200 samples per population, times two replicates for a total of ~1200 samples. These populations have been genotyped via GBS, and the Kresovich and Paterson lab groups are collecting important and extensive phenotypes on all these lines (e.g. stalk diameter, plant height, biomass yield, etc.). For this project, these populations can serve as 'validation' populations to confirm markers for various traits of interest including, but not limited to, the NIR panel of phenotypes.

Progress towards this aim

The NIR 'Forage' and 'Fuel' curves (Table A20) were developed by Chromatin using wet chemistry values obtained from Dairyland Laboratories on 150 diverse lines (50 sweet, 50 forage and 50 biomass). The 'NREL' curves were obtained from the National Renewable Energy Laboratory; these were developed based on wet chemistry data from the 150 diverse lines as well a proprietary panel developed by NREL. 1205 samples (~all 2012 experimental lines x 2 replicates)

and 3800 samples (~all 2013 experimental lines x 2 replicates) have been scanned, and values are now available for analysis. For this project, we will be focusing on compositional traits relevant to cellulosic biofuel and forage production including, but not limited to, lignin, NDFD, ash content, xylan and glucan.

Results and planned figures:

1) Identify means, ranges, difference within range for all traits
2) With multiple location/rep/year data collected, carry out analysis of variance (ANOVA) and calculate estimates of variance components (genotype and genotype x environment)
3) Calculate heritability for traits of interest (e.g. lignin, NDFD)
4) Assess associations among traits via PCA analysis
5) Carry out GWAS for compositional phenotypes to identify QTL(s)
6) Identify correlation between mineral phenotype and genetic structure
7) Calculate phenotypic correlations between the levels of various traits (e.g. lignin and NDFD)
8) Use multiple GWAS to identify traits with shared QTL

Expected results

Similar to the mineral QTL in seeds, it is possible that biomass composition QTL will also show pleiotropic effects on related traits. For example, there is a shared QTL in sugarcane for sugar yield, stalk weight and ash content [28]. In breeding, pleiotropic QTL can be utilized in selection practices (where selection can be based on correlated traits that are easy to select for). In a similar study in maize, QTL mapping in a RIL population revealed NDFD positively correlated to xylan

concentration and negatively correlated to lignin concentration. Glucan did not correlate to NDFD- a surprising result considering the presence of glucan in the crystalline microfibril structures of the cell wall [124]. We would expect to see similar correlations and trends in the compositional analysis of sorghum.

We generated and analyzed genetic and biomass NIR profiles for a large panel of sorghum lines, including ~400 genetically diverse sorghum lines and an additional 150 lines representing elite, bioenergy and sugar type sorghums. Samples were collected and analyzed in replicate over the course of the 2012 and 2013 field seasons in Florence, SC. We intend to map QTL in these populations for traits relevant to cellulosic biofuel production, including ash content, neutral detergent fiber digestibility (NDFD), glucan, xylan, and lignin concentration. The specific objectives of this study will be to 1) locate QTL for biomass traits with GWA analysis and 2) identify genotypic correlations, between-trait correlations and QTL colocation.

REFERENCES

1. Rosenow D, Quisenberry J, Wendt C, Clark L: **Drought tolerant sorghum and cotton germplasm**. *Agricultural Water Management* 1983, **7**(1):207-222.
2. Mann J, Kimber C, Miller F: **The origin and early cultivation of sorghums in Africa**. *Texas Agricultural Experiment Station Available electronically from* http://hdl *handle net/1969* 1983, **1**:128074.
3. de Vries SC, van de Ven GW, van Ittersum MK, Giller KE: **Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques**. *Biomass and Bioenergy* 2010, **34**(5):588-601.
4. Kimber CT, Dahlberg JA, Kresovich S: **The gene pool of Sorghum bicolor and its improvement**. In: *Genomics of the Saccharinae.* Springer; 2013: 23-41.
5. Shoemaker C, Bransby DI: **The role of sorghum as a bioenergy feedstock**. *Sustainable alternative fuel feedstock opportunities, challenges and roadmaps for six US regions Soil and Water Conserv Soc, Ankeny, IA* 2010:149-159.
6. Burow G, Burke JJ, Xin Z, Franks CD: **Genetic dissection of early-season cold tolerance in sorghum (Sorghum bicolor (L.) Moench)**. *Molecular Breeding* 2011, **28**(3):391-402.
7. Paterson AH, Bowers JE, Bruggmann Rm, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A: **The Sorghum bicolor genome and the diversification of grasses**. *Nature* 2009, **457**(7229):551-556.
8. Zhang J-Y, Lee Y-C, Torres-Jerez I, Wang M, Yin Y, Chou W-C, He J, Shen H, Srivastava AC, Pennacchio C: **Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (Panicum virgatum L.)**. *The Plant Journal* 2013.
9. Swaminathan K, Alabady MS, Varala K, De Paoli E, Ho I, Rokhsar DS, Arumuganathan AK, Ming R, Green PJ, Meyers BC: **Genomic and small RNA sequencing of Miscanthus x giganteus shows the utility of sorghum as a reference genome sequence for Andropogoneae grasses**. *Genome Biol* 2010, **11**(2):R12.
10. Welch RM, Graham RD: **Agriculture: the real nexus for enhancing bioavailable micronutrients in food crops**. *Journal of Trace Elements in Medicine and Biology* 2005, **18**(4):299-307.
11. Šimić D, Drinić SM, Zdunić Z, Jambrović A, Ledenčan T, Brkić J, Brkić A, Brkić I: **Quantitative trait loci for biofortification traits in maize grain**. *Journal of Heredity* 2012, **103**(1):47-54.

12. Malone JH, Oliver B: **Microarrays, deep sequencing and the true measure of the transcriptome**. *BMC biology* 2011, **9**(1):34.
13. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix GeneChip probe level data**. *Nucleic acids research* 2003, **31**(4):e15-e15.
14. Hamblin MT, Buckler ES, Jannink J-L: **Population genetics of genomics-based crop improvement methods**. *Trends in Genetics* 2011, **27**(3):98-106.
15. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES: **Genome-wide association study of leaf architecture in the maize nested association mapping population**. *Nature genetics* 2011, **43**(2):159-162.
16. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z: **Genome-wide association studies of 14 agronomic traits in rice landraces**. *Nature genetics* 2010, **42**(11):961-967.
17. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT: **Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines**. *Nature* 2010, **465**(7298):627-631.
18. Baxter IR, Ziegler G, Lahner B, Mickelbart MV, Foley R, Danku J, Armstrong P, Salt DE, Hoekenga OA: **Single-Kernel Ionomic Profiles Are Highly Heritable Indicators of Genetic and Environmental Influences on Elemental Accumulation in Maize Grain (Zea mays)**. *PLoS One* 2014, **9**(1):e87628.
19. Mamo BE, Barber BL, Steffenson BJ: **Genome-wide association mapping of zinc and iron concentration in barley landraces from Ethiopia and Eritrea**. *Journal of Cereal Science* 2014.
20. Nordborg M, Weigel D: **Next-generation genetics in plants**. *Nature* 2008, **456**(7223):720-723.
21. Bergelson J, Roux F: **Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana**. *Nature Reviews Genetics* 2010, **11**(12):867-879.
22. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nature genetics* 2005, **38**(2):203-208.
23. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M: **An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations**. *Nature genetics* 2012, **44**(7):825-830.
24. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E: **Efficient control of population structure in model organism association mapping**. *Genetics* 2008, **178**(3):1709-1723.
25. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE: **Population genomic and genome-wide association studies of agroclimatic traits in sorghum**. *Proceedings of the National Academy of Sciences* 2013, **110**(2):453-458.

26. Cardinal A, Lee M, Moore K: **Genetic mapping and analysis of quantitative trait loci affecting fiber and lignin content in maize**. *Theoretical and Applied Genetics* 2003, **106**(5):866-874.
27. Lung'aho MG, Mwaniki AM, Szalma SJ, Hart JJ, Rutzke MA, Kochian LV, Glahn RP, Hoekenga OA: **Genetic and physiological analysis of iron biofortification in maize kernels**. *PloS one* 2011, **6**(6):e20429.
28. Ming R, Wang Y, Draye X, Moore P, Irvine J, Paterson A: **Molecular dissection of complex traits in autopolyploids: mapping QTLs affecting sugar yield and related traits in sugarcane**. *Theoretical and Applied Genetics* 2002, **105**(2-3):332-345.
29. Stangoulis JC, Huynh B-L, Welch RM, Choi E-Y, Graham RD: **Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content**. *Euphytica* 2007, **154**(3):289-294.
30. Shakoor N, Nair R, Crasta O, Morris G, Feltus A, Kresovich S: **A Sorghum bicolor expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums**. *BMC plant biology* 2014, **14**(1):35.
31. Komili S, Silver PA: **Coupling and coordination in gene expression processes: a systems biology view**. *Nature Reviews Genetics* 2008, **9**(1):38-48.
32. Mane S, Evans C, Cooper K, Crasta O, Folkerts O, Hutchison S, Harkins T, Thierry-Mieg D, Thierry-Mieg J, Jensen R: **Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing**. *BMC genomics* 2009, **10**(1):264.
33. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N: **Phytozome: a comparative platform for green plant genomics**. *Nucleic acids research* 2012, **40**(D1):D1178-D1186.
34. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J: **Bioconductor: open software development for computational biology and bioinformatics**. *Genome biology* 2004, **5**(10):R80.
35. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: **NCBI GEO: mining tens of millions of expression profiles--database and tools update**. *Nucleic acids research* 2007, **35**(suppl 1):D760-D765.
36. Zimmermann P, Laule O, Schmitz J, Hruz T, Bleuler S, Gruissem W: **Genevestigator transcriptome meta-analysis and biomarker search using rice and barley gene expression databases**. *Molecular plant* 2008, **1**(5):851-857.
37. Joung J-G, Corbett AM, Fellman SM, Tieman DM, Klee HJ, Giovannoni JJ, Fei Z: **Plant MetGenMAP: an integrative analysis system for plant systems biology**. *Plant Physiology* 2009, **151**(4):1758-1768.
38. Katari MS, Nowicki SD, Aceituno FF, Nero D, Kelfer J, Thompson LP, Cabello JM, Davidson RS, Goldberg AP, Shasha DE: **VirtualPlant: a software platform to support systems biology research**. *Plant Physiology* 2010, **152**(2):500-515.

39. Buchanan CD, Lim S, Salzman RA, Kagiampakis I, Morishige DT, Weers BD, Klein RR, Pratt LH, Cordonnier-Pratt M-Ml, Klein PE: **Sorghum bicolor's transcriptome response to dehydration, high salinity and ABA**. *Plant Molecular Biology* 2005, **58**(5):699-720.
40. Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SÄ, Jiang N, Robin Buell C: **Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution**. *The Plant Journal* 2012, **71**(3):492-502.
41. Salzman RA, Brady JA, Finlayson SA, Buchanan CD, Summer EJ, Sun F, Klein PE, Klein RR, Pratt LH, Cordonnier-Pratt M-Ml: **Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses**. *Plant Physiology* 2005, **138**(1):352-368.
42. Jiang S-Y, Ma Z, Vanitha J, Ramachandran S: **Genetic variation and expression diversity between grain and sweet sorghum lines**. *BMC genomics* 2013, **14**(1):18.
43. Dugas DV, Monaco MK, Olson A, Klein RR, Kumari S, Ware D, Klein PE: **Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid**. *BMC genomics* 2011, **12**(1):514.
44. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU: **A gene expression map of Arabidopsis thaliana development**. *Nature genetics* 2005, **37**(5):501-506.
45. Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM: **Genome-wide atlas of transcription during maize development**. *The Plant Journal* 2011, **66**(4):553-563.
46. Druka A, Muehlbauer G, Druka I, Caldo R, Baumann U, Rostoks N, Schreiber A, Wise R, Close T, Kleinhofs A: **An atlas of gene expression from seed to seed through barley development**. *Functional & integrative genomics* 2006, **6**(3):202-211.
47. Jiao Y, Tausta SL, Gandotra N, Sun N, Liu T, Clay NK, Ceserani T, Chen M, Ma L, Holford M: **A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies**. *Nature genetics* 2009, **41**(2):258-263.
48. Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J: **A dynamic gene expression atlas covering the entire life cycle of rice**. *The Plant Journal* 2010, **61**(5):752-766.
49. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G: **An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants**. *The Plant Journal* 2010, **63**(1):86-99.
50. Das P, Samantaray S, Rout G: **Studies on cadmium toxicity in plants: a review**. *Environmental pollution* 1997, **98**(1):29-36.
51. Martin JH, Stephens JC: **The culture and use of sorghums for forage**: US Dept. of Agriculture; 1955.

52. Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW, Sasidharan R, Reinke V, Waterston RH, Gerstein M: **Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays**. *BMC genomics* 2010, **11**(1):383.
53. Kogenaru S, Yan Q, Guo Y, Wang N: **RNA-seq and microarray complement each other in transcriptome profiling**. *BMC genomics* 2012, **13**(1):629.
54. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays**. *Genome research* 2008, **18**(9):1509-1517.
55. Dallas PB, Gottardo NG, Firth MJ, Beesley AH, Hoffmann K, Terry PA, Freitas JR, Boag JM, Cummings AJ, Kees UR: **Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR - how well do they correlate?** *BMC genomics* 2005, **6**(1):59.
56. Morey JS, Ryan JC, Van Dolah FM: **Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR**. *Biological procedures online* 2006, **8**(1):175-193.
57. Ichihashi Y, Horiguchi G, Gleissberg S, Tsukaya H: **The bHLH transcription factor SPATULA controls final leaf size in Arabidopsis thaliana**. *Plant and cell physiology* 2010, **51**(2):252-261.
58. Lopez F, Bousser A, Sissoéff I, Hoarau J, Mahé A: **Characterization in maize of ZmTIP2-3, a root-specific tonoplast intrinsic protein exhibiting aquaporin activity**. *Journal of experimental botany* 2004, **55**(396):539-541.
59. Meyer E, Logan TL, Juenger TE: **Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research**. *The Plant Journal* 2012, **70**(5):879-890.
60. Du Z, Zhou X, Ling Y, Zhang Z, Su Z: **agriGO: a GO analysis toolkit for the agricultural community**. *Nucleic acids research* 2010, **38**(suppl 2):W64-W70.
61. Qazi HA, Paranjpe S, Bhargava S: **Stem sugar accumulation in sweet sorghum - activity and expression of sucrose metabolizing enzymes and sucrose transporters**. *Journal of plant physiology* 2012, **169**(6):605-613.
62. Zhao Q, Tobimatsu Y, Zhou R, Pattathil S, Gallego-Giraldo L, Fu C, Jackson LA, Hahn MG, Kim H, Chen F: **Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in Medicago truncatula**. *Proceedings of the National Academy of Sciences* 2013, **110**(33):13660-13665.
63. Saballos A, Sattler SE, Sanchez E, Foster TP, Xin Z, Kang C, Pedersen JF, Vermerris W: **Brown midrib2 (Bmr2) encodes the major 4‐coumarate: coenzyme A ligase involved in lignin biosynthesis in sorghum (Sorghum bicolor (L.) Moench)**. *The Plant Journal* 2012, **70**(5):818-830.
64. Jouanin L, Lapierre C: **Lignins: Biosynthesis, Biodegradation and Bioengineering**: Elsevier Science; 2012.
65. Vanderlip R, Reeves H: **Growth stages of sorghum [Sorghum bicolor,(L.) Moench.]**. *Agronomy Journal* 1972, **64**(1):13-16.

66. Whitfield MB, Chinn MS, Veal MW: **Processing of materials derived from sweet sorghum for biobased products**. *Industrial Crops and Products* 2012, **37**(1):362-375.
67. Livak KJ, Schmittgen TD: **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2< sup>,àí ŒîŒîCT</sup> Method**. *methods* 2001, **25**(4):402-408.
68. Amaducci S, Monti A, Venturi G: **Non-structural carbohydrates and fibre components in sweet and fibre sorghum as affected by low and normal input techniques**. *Industrial Crops and Products* 2004, **20**(1):111-118.
69. Obernberger I, Biedermann F, Widmann W, Riedl R: **Concentrations of inorganic elements in biomass fuels and recovery in the different ash fractions**. *Biomass and bioenergy* 1997, **12**(3):211-224.
70. Monti A, Di Virgilio N, Venturi G: **Mineral composition and ash content of six major energy crops**. *Biomass and Bioenergy* 2008, **32**(3):216-223.
71. Organization WH: **The world health report 2002: reducing risks, promoting healthy life**: World Health Organization; 2002.
72. Graham R, Senadhira D, Beebe S, Iglesias C, Monasterio I: **Breeding for micronutrient density in edible portions of staple food crops: conventional approaches**. *Field Crops Research* 1999, **60**(1):57-80.
73. White PJ, Broadley MR: **Biofortifying crops with essential mineral elements**. *Trends in plant science* 2005, **10**(12):586-593.
74. Ma JF, Yamaji N, Mitani N, Xu X-Y, Su Y-H, McGrath SP, Zhao F-J: **Transporters of arsenite in rice and their role in arsenic accumulation in rice grain**. *Proceedings of the National Academy of Sciences* 2008, **105**(29):9931-9935.
75. Vreugdenhil D, Aarts M, Koornneef M, Nelissen H, Ernst W: **Natural variation and QTL analysis for cationic mineral content in seeds of Arabidopsis thaliana**. *Plant, Cell & Environment* 2004, **27**(7):828-839.
76. Yi Y, Guerinot ML: **Genetic evidence that induction of root Fe (III) chelate reductase activity is necessary for iron uptake under iron deficiency†**. *The Plant Journal* 1996, **10**(5):835-844.
77. Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot ML, Briat J-F, Curie C: **IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth**. *The Plant Cell Online* 2002, **14**(6):1223-1233.
78. Connolly EL, Campbell NH, Grotz N, Prichard CL, Guerinot ML: **Overexpression of the FRO2 ferric chelate reductase confers tolerance to growth on low iron and uncovers posttranscriptional control**. *Plant Physiology* 2003, **133**(3):1102-1110.
79. Baxter IR, Vitek O, Lahner B, Muthukumar B, Borghi M, Morrissey J, Guerinot ML, Salt DE: **The leaf ionome as a multivariable system to detect a plant's physiological status**. *Proceedings of the National Academy of Sciences* 2008, **105**(33):12081-12086.
80. Ortiz-Monasterio J, Palacios-Rojas N, Meng E, Pixley K, Trethowan R, Pena R: **Enhancing the mineral and vitamin content of wheat and maize through plant breeding**. *Journal of Cereal Science* 2007, **46**(3):293-307.

81. Waters BM, Grusak MA: **Quantitative trait locus mapping for seed mineral concentrations in two Arabidopsis thaliana recombinant inbred populations**. *New Phytologist* 2008, **179**(4):1033-1047.
82. Norton GJ, Deacon CM, Xiong L, Huang S, Meharg AA, Price AH: **Genetic mapping of the rice ionome in leaves and grain: identification of QTLs for 17 elements including arsenic, cadmium, iron and selenium**. *Plant and soil* 2010, **329**(1-2):139-153.
83. Shi R, Li H, Tong Y, Jing R, Zhang F, Zou C: **Identification of quantitative trait locus of zinc and phosphorus density in wheat (Triticum aestivum L.) grain**. *Plant and soil* 2008, **306**(1-2):95-104.
84. Peleg Z, Cakmak I, Ozturk L, Yazici A, Jun Y, Budak H, Korol AB, Fahima T, Saranga Y: **Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat× wild emmer wheat RIL population**. *Theoretical and Applied Genetics* 2009, **119**(2):353-369.
85. Casa AM, Pressoir G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S: **Community resources and strategies for association mapping in sorghum**. *Crop science* 2008, **48**(1):30-40.
86. Lestienne I, Icard-Vernière C, Mouquet C, Picq C, Trèche S: **Effects of soaking whole cereal and legume seeds on iron, zinc and phytate contents**. *Food Chemistry* 2005, **89**(3):421-425.
87. Mengesha MH: **Chemical composition of teff (Eragrostis tef) compared with that of wheat, barley and grain sorghum**. *Economic Botany* 1966, **20**(3):268-273.
88. Ragaee S, Abdel-Aal E-SM, Noaman M: **Antioxidant activity and nutrient composition of selected cereals for food use**. *Food Chemistry* 2006, **98**(1):32-38.
89. Neucere NJ, Sumrell G: **Chemical composition of different varieties of grain sorghum**. *Journal of agricultural and food chemistry* 1980, **28**(1):19-21.
90. Gomez-Becerra HF, Yazici A, Ozturk L, Budak H, Peleg Z, Morgounov A, Fahima T, Saranga Y, Cakmak I: **Genetic variation and environmental stability of grain mineral nutrient concentrations in Triticum dicoccoides under five environments**. *Euphytica* 2010, **171**(1):39-52.
91. Ozgen S, Busse JS, Palta JP: **Influence of Root Zone Calcium on Shoot Tip Necrosis and Apical Dominance of Potato Shoot: Simulation of This Disorder by Ethylene Glycol Tetra Acetic Acid and Prevention by Strontium**. *HortScience* 2011, **46**(10):1358-1362.
92. Queen WH, Fleming HW, O'Kelley JC: **Effects on Zea mays seedlings of a strontium replacement for calcium in nutrient media**. *Plant physiology* 1963, **38**(4):410.
93. Hutchin ME, Vaughan BE: **Relation between simultaneous Ca and Sr transport rates in isolated segments of vetch, barley, and pine roots**. *Plant physiology* 1968, **43**(12):1913-1918.
94. Broadley MR, White PJ: **Some elements are more equal than others: soil-to-plant transfer of radiocaesium and radiostrontium, revisited**. *Plant and soil* 2012, **355**(1):23-27.

95. Maathuis FJ: **Physiological functions of mineral macronutrients**. *Current opinion in plant biology* 2009, **12**(3):250-258.
96. Marschner H, Marschner P: **Marschner's mineral nutrition of higher plants**, vol. 89: Academic press; 2012.
97. Conover WJ, Iman RL: **Rank transformations as a bridge between parametric and nonparametric statistics**. *The American Statistician* 1981, **35**(3):124-129.
98. Zhang T, Zhao X, Wang W, Pan Y, Huang L, Liu X, Zong Y, Zhu L, Yang D, Fu B: **Comparative transcriptome profiling of chilling stress responsiveness in two contrasting rice genotypes**. *PloS one* 2012, **7**(8):e43274.
99. Gregorio GB, Senadhira D, Htut H, Graham RD: **Breeding for trace mineral density in rice**. *Food & Nutrition Bulletin* 2000, **21**(4):382-386.
100. Salt DE, Baxter I, Lahner B: **Ionomics and the study of the plant ionome**. *Annu Rev Plant Biol* 2008, **59**:709-733.
101. Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M: **Genome Wide Association in tomato reveals 44 candidate loci for fruit metabolic traits**. *Plant Physiology* 2014:pp. 114.241521.
102. Zhao K, Tung C-W, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J: **Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa**. *Nature communications* 2011, **2**:467.
103. Aschner JL, Aschner M: **Nutritional aspects of manganese homeostasis**. *Molecular aspects of medicine* 2005, **26**(4):353-362.
104. Delhaize E, Gruber BD, Pittman JK, White RG, Leung H, Miao Y, Jiang L, Ryan PR, Richardson AE: **A role for the AtMTP11 gene of Arabidopsis in manganese transport and tolerance**. *The Plant Journal* 2007, **51**(2):198-210.
105. Inaba T, Kobayashi E, Suwazono Y, Uetani M, Oishi M, Nakagawa H, Nogawa K: **Estimation of cumulative cadmium intake causing Itai–itai disease**. *Toxicology letters* 2005, **159**(2):192-201.
106. Arao T, Ae N: **Genotypic variations in cadmium levels of rice grain**. *Soil Science and Plant Nutrition* 2003, **49**(4):473-479.
107. Chao D-Y, Silva A, Baxter I, Huang YS, Nordborg M, Danku J, Lahner B, Yakubova E, Salt DE: **Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in Arabidopsis thaliana**. *PLoS genetics* 2012, **8**(9):e1002923.
108. Yusuf M, Fariduddin Q, Hayat S, Ahmad A: **Nickel: an overview of uptake, essentiality and toxicity in plants**. *Bulletin of environmental contamination and toxicology* 2011, **86**(1):1-17.
109. Curie C, Cassin G, Couch D, Divol F, Higuchi K, Le Jean M, Misson J, Schikora A, Czernic P, Mari S: **Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters**. *Annals of botany* 2009, **103**(1):1-11.
110. Curie C, Panaviene Z, Loulergue C, Dellaporta SL, Briat J-F, Walker EL: **Maize yellow stripe1 encodes a membrane protein directly involved in Fe (III) uptake**. *Nature* 2001, **409**(6818):346-349.

111. Gendre D, Czernic P, Conéjéro G, Pianelli K, Briat JF, Lebrun M, Mari S: **TcYSL3, a member of the YSL gene family from the hyper‐accumulator Thlaspi caerulescens, encodes a nicotianamine‐Ni/Fe transporter**. *The Plant Journal* 2007, **49**(1):1-15.
112. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J, Zhao K, Lee J, Hawkesford MJ, Guerinot ML: **Variation in molybdenum content across broadly distributed populations of Arabidopsis thaliana is controlled by a mitochondrial molybdenum transporter (MOT1)**. *PLoS Genetics* 2008, **4**(2):e1000004.
113. Ascher J, Graham R, Elliott D, Scott J, Jessop R: **Agronomic value of seed with high nutrient content**. *Wheat in heat stressed environments: irrigated dry areas and rice wheat farming systems Saunders & Hettel editores Mexico DF CIMMYT* 1994:297-308.
114. Modi AT: **Wheat seed quality in response to molybdenum and phosphorus**. *Journal of plant nutrition* 2002, **25**(11):2409-2419.
115. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E, Li Y, Bergelson J, Borevitz JO, Nordborg M: **A coastal cline in sodium accumulation in Arabidopsis thaliana is driven by natural variation of the sodium transporter AtHKT1; 1**. *PLoS genetics* 2010, **6**(11):e1001193.
116. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species**. *PloS one* 2011, **6**(5):e19379.
117. Box GE, Cox DR: **An analysis of transformations**. *Journal of the Royal Statistical Society Series B (Methodological)* 1964:211-252.
118. Team RC: **R: A language and environment for statistical computing**. 2012.
119. Venables WN, Ripley BD: **Modern applied statistics with S**: Springer; 2002.
120. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: **GAPIT: genome association and prediction integrated tool**. *Bioinformatics* 2012, **28**(18):2397-2399.
121. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *Journal of the Royal Statistical Society Series B (Methodological)* 1995:289-300.
122. Pan W, Lin J, Le CT: **How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach**. *Genome Biol* 2002, **3**(5):1-0022.0010.
123. Goodwin RA: **Soil Survey of Craven County, North Carolina**: The Service; 1989.
124. Lorenz A, Coors J, Hansey C, Kaeppler S, De Leon N: **Genetic analysis of cell wall traits relevant to cellulosic ethanol production in maize (L.)**. *Crop Science* 2010, **50**(3):842-852.

APPENDIX A

SUPPLEMENTARY TABLES

Table A.1 Sorghum samples included in the gene expression atlas

Major organ	Tissue location	Tissue section	Rep	DAP	Genotype	Itype	Designation
leaf	leaf	all	1	38	PI455230	BTU	leaf_38DAP_ PI455230
leaf	leaf	all	1	38	R159	grain	leaf_38DAP_ R159
leaf	leaf	all	1	38	Atlas	sweet	leaf_38_Atlas
leaf	leaf	all	1	38	PI152611	forage	leaf_38_PI152611
leaf	leaf	all	1	38	AR2400	BTU	leaf_38_AR2400
leaf	leaf	all	1	38	Fremont	sweet	leaf_38_Fremont
seedling	root	all	1	10	PI455230	BTU	seedling_root_PI455230
seedling	root	all	1	10	R159	grain	seedling_root_R159
seedling	root	all	1	10	Atlas	sweet	seedling_root_Atlas
seedling	root	all	1	10	PI152611	forage	seedling_root_PI152611
seedling	root	all	1	10	AR2400	BTU	seedling_root_AR2400_1
seedling	root	all	2	10	AR2400	BTU	seedling_root_AR2400_2
seedling	root	all	1	10	Fremont	sweet	seedling_root_Fremont_1
seedling	root	all	2	10	Fremont	sweet	seedling_root_Fremont_2
seedling	shoot	all	1	10	PI455230	BTU	seedling_shoot_PI455230_1
seedling	shoot	all	2	10	PI455230	BTU	seedling_shoot_PI455230_2
seedling	shoot	all	1	10	R159	grain	seedling_shoot_R159_1
seedling	shoot	all	2	10	R159	grain	seedling_shoot_R159_2
seedling	shoot	all	1	10	Atlas	sweet	seedling_shoot_Atlas_1
seedling	shoot	all	2	10	Atlas	sweet	seedling_shoot_Atlas_2
seedling	shoot	all	1	10	PI152611	forage	seedling_shoot_PI152611_1
seedling	shoot	all	2	10	PI152611	forage	seedling_shoot_PI152611_2
seedling	shoot	all	1	10	AR2400	BTU	seedling_shoot_AR2400_1
seedling	shoot	all	2	10	AR2400	BTU	seedling_shoot_AR2400_2
seedling	shoot	all	1	10	Fremont	sweet	seedling_shoot_Fremont_1
seedling	shoot	all	2	10	Fremont	sweet	seedling_shoot_Fremont_2
shoot	shoot_tip	all	1	38	PI455230	BTU	shoot_tip_38DAP_PI455230
shoot	shoot_tip	all	1	38	R159	grain	shoot_tip_38DAP_ R159
shoot	shoot_tip	all	1	38	Atlas	sweet	shoot_tip_38_Atlas
shoot	shoot_tip	all	1	38	PI152611	forage	shoot_tip_38_PI152611

Table A.1 (cont.) Sorghum samples included in the gene expression atlas

Major organ	Tissue location	Tissue section	Rep	DAP	Genotype	Itype	Designation
shoot	shoot_tip	all	1	38	AR2400	BTU	shoot_tip_38_AR2400
shoot	shoot_tip	all	1	38	Fremont	sweet	shoot_tip_38_Fremont
stem	internode	top	1	61	PI455230	BTU	internode_61DAP_PI455230
stem	internode	top	1	61	R159	grain	internode_61DAP_R159
stem	internode	middle	1	70	PI455230	BTU	internode_70DAP_PI455230
stem	internode	bottom	1	61	PI455230	BTU	internode_61DAP_PI455230
stem	internode	bottom	1	61	R159	grain	internode_61DAP_R159
stem	internode	top	1	61	Atlas	sweet	internode_61DAP_Atlas
stem	internode	middle	1	70	Atlas	sweet	internode_70DAP_Atlas
stem	internode	bottom	1	61	Atlas	sweet	internode_61DAP_Atlas
stem	internode	top	1	61	PI152611	forage	internode_61DAP_PI152611
stem	internode	middle	1	70	PI152611	forage	internode_70DAP_PI152611
stem	internode	bottom	1	61	PI152611	forage	internode_61DAP_PI152611
stem	internode	top	1	61	AR2400	BTU	internode_61DAP_AR2400
stem	internode	middle	1	70	AR2400	BTU	internode_70DAP_AR2400
stem	internode	bottom	1	61	AR2400	BTU	internode_61DAP_AR2400
stem	internode	top	1	61	Fremont	sweet	internode_61DAP_Fremont
stem	internode	middle	1	70	Fremont	sweet	internode_70DAP_Fremont
stem	internode	bottom	1	61	Fremont	sweet	internode_61DAP_Fremont
stem	pith	top	1	70	R159	grain	pith_70DAP_R159
stem	pith	middle	1	70	PI455230	BTU	pith_70DAP_PI455230
stem	pith	bottom	1	70	PI455230	BTU	pith_70DAP_PI455230
stem	pith	all	1	60	PI455230	BTU	pith_60DAP_PI455230_1
stem	pith	all	2	60	PI455230	BTU	pith_60DAP_PI455230_2
stem	pith	all	3	60	PI455230	BTU	pith_60DAP_PI455230_3
stem	pith	bottom	1	70	R159	grain	pith_70DAP_R159_B
stem	pith	all	1	60	R159	grain	pith_60DAP_R159
stem	pith	top	1	70	Atlas	sweet	pith_70DAP_Atlas
stem	pith	middle	1	70	Atlas	sweet	pith_70DAP_Atlas
stem	pith	bottom	1	70	Atlas	sweet	pith_70DAP_Atlas
stem	pith	all	1	60	Atlas	sweet	pith_60DAP_Atlas_1
stem	pith	all	2	60	Atlas	sweet	pith_60DAP_Atlas_2
stem	pith	all	3	60	Atlas	sweet	pith_60DAP_Atlas_3
stem	rind	top	1	70	PI455230	BTU	rind_70DAP_PI455230
stem	rind	top	1	70	R159	grain	rind_70DAP_R159
stem	rind	middle	1	70	PI455230	BTU	rind_70DAP_PI455230
stem	rind	bottom	1	70	PI455230	BTU	rind_70DAP_PI455230
stem	rind	all	1	60	PI455230	BTU	rind_60DAP_PI455230_1

Table A.1 (cont.) Sorghum samples included in the gene expression atlas

Major organ	Tissue location	Tissue section	Rep	DAP	Genotype	Itype	Designation
stem	rind	all	2	60	PI455230	BTU	rind_60DAP_PI455230_2
stem	rind	all	3	60	PI455230	BTU	rind_60DAP_PI455230_3
stem	rind	all	1	60	R159	grain	rind_60DAP_R159
stem	rind	top	1	70	Atlas	sweet	rind_70DAP_Atlas
stem	rind	middle	1	70	Atlas	sweet	rind_70DAP_Atlas
stem	rind	bottom	1	70	Atlas	sweet	rind_70DAP_Atlas
stem	rind	all	1	60	Atlas	sweet	rind_60DAP_Atlas_1
stem	rind	all	2	60	Atlas	sweet	rind_60DAP_Atlas_2
stem	rind	all	3	60	Atlas	sweet	rind_60DAP_Atlas_3

Table A.2 Phenotypic characteristics of sorghum genotypes included in gene expression atlas

Trait	R159	Atlas	Fremont	PI152611	AR2400	PI455230
Ideotype	Grain	Sweet	Sweet	Biomass/ Forage	Biomass	Biomass
Plant Height (cm to top of head at mid bloom	92	185-320	171	410-420	153-178	285-350
Maturity (days to mid-bloom)	68	56-67	48	95	52-85	65
Grain Weight	2.7-3.2	1.7-2.8	1.6	4.6-4.9	2.6-2.7	2.1
Brix	NA	14.8	9.3	NA	NA	NA
Sucrose	NA	9.75	3.45	NA	NA	NA

* Data obtained from GRIN, National Genetic Resources Program (http://www.ars-grin.gov/)

** Brix value for Atlas genotype obtained from Kawahigashi, Hiroyuki, et al. "Evaluation of Brix and sugar content in stem juice from sorghum varieties." Grassland Science (2013).

Table A.3 Pearson's correlation coefficient of the biological replicates

Tissue	Genotype	Biological Rep 1vs2	P-value
shoot	PI455230	0.988	< 2.2e-16
shoot	R159	0.991	< 2.2e-16
shoot	Atlas	0.990	< 2.2e-16
shoot	PI152611	0.991	< 2.2e-16
shoot	AR2400	0.989	< 2.2e-16
shoot	Fremont	0.990	< 2.2e-16
root	AR2400	0.988	< 2.2e-16
root	Fremont	0.992	< 2.2e-16

* Calculated using cor.test in R

Table A.4 Expression of sorghum homologs with established patterns of expression in related species

Gene ID	Description	Previously-identified expression pattern	Study	Internode	Seedling Shoot	Seedling Root	Leaf	Shoot Tip	Pith	Rind
Sb04g036450	similar to Putative SPATULA	SPATULA expression highest in shoot tips and young leaf primordia	Ichihashi et al. 2010	26.50	**89.77**	31.22	45.47	**80.45**	13.85	36.40
Sb06g024590	similar to Aquaporin TIP2-3	Root specific expression	Lopez et al. 2004	114.75	246.56	**9504.49**	368.45	2056.38	1486.68	319.36

*Descriptions obtained from Phytozome (http://www.phytozome.net)

Table A.5 Gene Ontology classifications in the biological processes category identified using the AgriGO Singular Enrichment Analysis (SEA)

Biological Process	GO annotation	Shoot	Root	Leaf	Shoot tip	Stem	Constitutive Genes
Development							
GO: 0003006	Reproductive developmental process	772	791	737	848	823	265
GO: 0048856	Anatomical structure development	1591	1680	1484	1756	1654	350
GO: 0009791	Post-embryonic development	1023	1049	978	1107	1072	482
GO: 0009790	Embryonic development	803	857	782	912	895	413
GO: 0030154	Cell differentiation	512	541	0	579	543	235
GO: 0090066	Regulation of anatomical structure size	0	284	0	279	0	0
Metabolism							
GO: 0019725	Cellular homeostasis	211	207	229	217	205	95
GO: 0006091	Generation of precursor metabolites and energy	320	271	323	321	311	157
GO: 0006519	Cellular amino acid and derivative metabolic process	525	531	549	492	501	0
GO: 0006139	Nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process	0	0	0	2295	2246	0
Gene expression							
GO: 0010467	Gene expression	1987	2036	1930	2139	2162	912
Signaling/Transport							
GO: 0007267	Cell-cell signaling	64	66	0	71	66	0
GO: 0006810	Transport	1627	1671	1550	1639	1630	0
GO: 0007165	Signal transduction	0	0	0	0	0	377

Table A.5 (cont.) Gene Ontology classifications in the biological processes category identified using the AgriGO Singular Enrichment Analysis (SEA)

Biological Process	GO annotation	Shoot	Root	Leaf	Shoot tip	Stem	Constitutive Genes
Synthesis/metabolism							
GO:0006454	Protein modification process	1038	1098	0	1104	1036	456
GO: 0006412	Translation	514	479	501	505	509	285
GO: 0019538	Protein metabolic process	2134	2135	2060	2200	2105	991
Carbohydrate synthesis/metabolism							
GO: 0005975	Carbohydrate metabolic process	647	640	600	679	592	268
Lipid synthesis/metabolism							
GO: 0006629	Lipid metabolic process	616	585	577	568	0	0
Response to stimulus							
GO: 0050896	Response to stimulus	2516	2559	2430	2501	2430	1038
Photosynthesis							
GO: 0015979	Photosynthesis	143	0	148	0	120	0
No annotation	No annotation	1672	1558	1687	1611	1671	626
Total # of annotated genes		11425	11338	11123	11484	11247	4526

* GO annotations were obtained from agriGO (Du et al., 2010; http://bioinfo.cau.edu.cn/agriGO/) for the *Sorghum bicolor* gene models

** Statistical significance was detected with the Fisher's exact test (p-value ≤0.05) and the Yekutieli (false-discovery rate under dependency) multi-test adjustment method

Table A.6 Identification of stably expressed genes (CV </= 15%). Please note that this is an excerpt of Table A6. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

Gene ID	Mean	%CV	seedling_shoot _PI455230_1	seedling_shoot _PI455230_2	seedling_root _PI455230	leaf_38DAP _ PI455230	shoot_tip_38DAP _ PI455230
Sb05g016480	21888.21	5.18	19155.78	20393.50	23279.62	19547.81	22043.93
Sb03g039445	13846.23	6.62	14071.31	13872.44	11843.14	13349.36	14138.03
Sb05g016450	13417.94	7.47	12865.19	12050.47	13152.75	11712.20	13826.66
Sb03g030970	6933.11	8.45	6778.36	6661.22	7415.57	5975.49	8397.05
Sb02g000760	424.02	9.15	372.15	437.64	438.01	429.56	440.85
Sb06g002790	902.65	9.26	826.48	915.36	918.37	849.70	958.36
Sb06g017780	771.30	9.38	787.80	695.16	787.53	865.56	703.58
Sb02g024290	1524.85	9.42	1710.76	1434.91	1546.42	1490.36	1617.34
Sb05g005380	659.34	9.96	606.31	630.62	671.70	642.06	668.51
Sb09g023560*	8074.26	10.38	7611.84	7424.88	7776.28	6448.74	7867.23
Sb10g008050	705.38	10.39	735.94	663.03	712.05	599.82	707.69
Sb07g002220	9419.39	10.48	9702.17	8796.34	11400.70	7498.06	10766.64
Sb05g004130	1919.65	10.48	1891.13	2080.42	2359.49	1799.78	2577.09
Sb04g006570	441.03	10.49	489.48	393.02	435.67	538.24	420.44
Sb03g038120	8114.70	10.51	7621.98	7362.00	8504.92	6399.55	8048.30
Sb01g004540	648.23	10.52	712.38	584.87	695.27	604.50	684.58
Sb01g040220	688.94	10.54	594.37	672.03	667.18	628.57	755.81
Sb09g028000	533.64	10.57	500.48	477.74	492.39	574.94	583.52
Sb04g023310	1067.53	10.60	1052.21	860.63	1018.25	1176.02	1196.34
Sb03g008550	9370.60	10.63	8344.72	7820.61	10271.59	6994.57	8554.14
Sb09g024010	439.99	10.64	433.40	432.81	332.68	459.66	473.98
Sb09g000610	1618.87	10.68	1489.27	1524.60	1652.59	1487.45	1518.77

Table A.7 List of small RNAs (sorghum, maize and sugarcane) included in microarray design. Please note that this is an excerpt of Table A7. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

PMRD_Sorghum bicolor (Sorghum) miRNA	PMRD_Sorghum bicolor (Sorghum) miRNA	PMRD_Zea mays (Corn) miRNA	PMRD_Zea mays (Maize) miRNA	PMRD_Saccharum (Sugarcane) miRNA	PMRD_Saccharum (Sugarcane) miRNA
Stem-loop sequences	*Mature sequences*	*Stem-loop sequences*	*Mature sequences*	*Stem-loop sequences*	*Mature sequences*
sbi-MIR156a	sbi-miR156a	zma-MIR156a	zma-miR156a	ssp-MIR167	ssp-miR167
sbi-MIR156b	sbi-miR156b	zma-MIR156b	zma-miR156b	ssp-MIR168	ssp-miR168
sbi-MIR156c	sbi-miR156c	zma-MIR156c	zma-miR156c		
sbi-MIR156d	sbi-miR156d	zma-MIR156d	zma-miR156d		
sbi-MIR156e	sbi-miR156e	zma-MIR156e	zma-miR156e		
sbi-MIR156f	sbi-miR156f	zma-MIR156f	zma-miR156f		
sbi-MIR156g	sbi-miR156g	zma-MIR156g	zma-miR156g		
sbi-MIR156h	sbi-miR156h	zma-MIR156h	zma-miR156h		
sbi-MIR156i	sbi-miR156i	zma-MIR156i	zma-miR156i		
sbi-MIR157	sbi-miR157	zma-MIR156j	zma-miR156j		
sbi-MIR159	sbi-miR159	zma-MIR156k	zma-miR156k		
sbi-MIR159b	sbi-miR159b	zma-MIR156l	zma-miR156l		
sbi-MIR160a	sbi-miR160a	zma-MIR156l	zma-miR156l*		
sbi-MIR160b	sbi-miR160b	zma-MIR156m	zma-miR156m		
sbi-MIR160c	sbi-miR160c	zma-MIR156n	zma-miR156n		
sbi-MIR160d	sbi-miR160d	zma-MIR156o	zma-miR156o		
sbi-MIR160e	sbi-miR160e	zma-MIR156p	zma-miR156p		
sbi-MIR160f	sbi-miR160f	zma-MIR156q	zma-miR156q		
sbi-MIR162	sbi-miR162	zma-MIR156r	zma-miR156r		
sbi-MIR164	sbi-miR164	zma-MIR157m	zma-miR157m		

** PMRD sequences were obtained at the plant microRNA database http://bioinformatics.cau.edu.cn/PMRD/*

Table A.7 (cont.) List of small RNAs (sorghum, maize and sugarcane) included in microarray design. Please note that this is an excerpt of Table A7. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

fRNAdb (Sorghum bicolor)	fRNAdb (Sorghum bicolor)	fRNAdb (Zea mays)	fRNAdb (Zea mays)	fRNAdb (Sugarcane)	fRNAdb (Sugarcane)
ID	Description	*ID*	Description	*ID*	Description
FR401736	5.8S ribosomal RNA (rRNA)	FR299237	5.8S ribosomal RNA (rRNA)	FR055686	Group II intron
FR126945	precursor micro RNA (miRNA) mir-164	FR282092	Group II intron	FR111223	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR092110	C/D box guide small nucleolar RNA (snoRNA) R12	FR301119	Group II intron	FR244794	transfer RNA (tRNA), GGT (Thr/T) Threonine
FR024455	precursor micro RNA (miRNA) mir-172c	FR234963	transfer RNA (tRNA), CAT (Met/M) Methionine	FR103328	transfer RNA (tRNA), GAA (Phe/F) Phenylalanine
FR060613	precursor micro RNA (miRNA) mir-156c	FR077192	transfer RNA (tRNA), CAT (Met/M) Methionine	FR043851	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR084780	precursor micro RNA (miRNA) mir-160b	FR202821	5.8S ribosomal RNA (rRNA)	FR313592	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR087762	precursor micro RNA (miRNA) mir-172a	FR378790	transfer RNA (tRNA), TGA (Ser/S) Serine	FR334481	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR091580	precursor micro RNA (miRNA) mir-160e	FR009172	5.8S ribosomal RNA (rRNA)	FR360253	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR194965	precursor micro RNA (miRNA) mir-172b	FR022869	5.8S ribosomal RNA (rRNA)	FR361165	transfer RNA (tRNA), GAT (Ile/I) Isoleucine
FR203951	precursor micro RNA (miRNA) mir-160a	FR035707	5.8S ribosomal RNA (rRNA)	FR101425	transfer RNA (tRNA), TTG (Gln/Q) Glutamine
FR216966	precursor micro RNA (miRNA) mir-166d	FR096962	5.8S ribosomal RNA (rRNA)	FR286796	transfer RNA (tRNA), GAA (Phe/F) Phenylalanine
FR246000	precursor micro RNA (miRNA) mir-167b	FR112226	5.8S ribosomal RNA (rRNA)	FR323811	transfer RNA (tRNA), CAT (Met/M) Methionine
FR251647	precursor micro RNA (miRNA) mir-169a	FR159491	5.8S ribosomal RNA (rRNA)	FR369828	transfer RNA (tRNA), TCT (Arg/R) Arginine

** fRNAdb sequences were obtained at the functional noncoding RNA sequence database http://www.ncrna.org/frnadb/*

Table A.8 Expression of select sucrose metabolizing enzyme/transporter genes. Please note that this is an excerpt of Table A8. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

Sucrose transporter genes	Probes	seedling_shoot _PI455230_1	seedling_shoot _PI455230_2	seedling_root _PI455230	leaf_38DAP _ PI455230	shoottip_38DAP _ PI455230
SUT1	Sb01g045720	11.77	11.78	11.90	12.16	9.51
SUT2	Sb04g038030	9.86	9.02	9.54	9.21	9.45
SUT3	Sb01g022430	5.04	5.09	4.95	5.77	5.29
SUT4	Sb08g023310	9.29	8.62	8.47	9.89	8.63
SUT5	Sb04g023860	7.45	6.66	6.73	8.09	7.83
Invertase genes						
INV1 (cell wall)	Sb01g008910	6.21	5.48	6.41	6.39	6.08
INV2 (cell wall)	Sb03g047060	6.77	6.73	8.91	5.65	5.82
INV3 (vacuolar)	Sb04g000620	11.03	11.11	12.59	12.20	11.35
INV4 (cell wall)	Sb06g031930	7.39	7.32	8.63	7.44	7.43
Sucrose phosphate synthase genes						
SPS1	Sb03g043900	10.36	10.67	6.55	11.65	6.37
SPS2	Sb04g005720	10.34	10.73	10.69	11.52	10.49
SPS3	Sb05g007310	9.83	9.97	7.69	11.51	7.37
SPS4	Sb09g028570	9.93	10.03	10.62	9.17	10.49
SPS5	Sb10g025240	10.54	10.92	10.68	11.50	10.60
Sucrose synthase genes						
SUS1	Sb01g035890	5.80	6.04	7.66	7.22	7.12
SUS2	Sb04g038410	6.86	7.09	7.06	5.58	8.32
SUS3	Sb10g031040	12.08	12.17	11.61	8.14	12.80
Constitutively expressed gene						
Ubiquitin	Sb10g027470	11.71	11.47	11.74	11.22	11.05
eIF4A.1	Sb04g003390	11.50	11.33	12.36	10.33	12.42

* RMA-normalized log2 expression values

Table A.9 Expression of select phenylpropanoid-monolignol biosynthesis pathway genes. Please note that this is an excerpt of Table A9. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

Phenylalanine ammonia lyase (PAL)	Probes	seedling_shoot _PI455230_1	seedling_shoot _PI455230_2	seedling_root _PI455230	leaf_38DAP _ PI455230	shoot_tip_38DAP _ PI455230
PAL	Sb01g014020	12.80	12.80	13.46	12.74	12.72
PAL	Sb04g026510	11.05	10.75	11.86	10.63	10.55
PAL	Sb04g026520	9.62	9.95	10.77	9.88	9.95
PAL	Sb04g026530	9.48	9.12	10.70	9.96	10.06
PAL	Sb04g026540	9.41	9.13	10.70	10.01	9.89
PAL	Sb04g026550	9.46	9.12	10.67	9.85	9.96
PAL	Sb04g026560	9.46	9.25	10.68	9.99	10.06
PAL	Sb06g022740	11.03	10.61	11.84	10.56	10.45
PAL	Sb06g022750	12.92	12.71	13.35	12.77	12.74
p-coumarate 3-hydroxylase (C3H)						
Sb09g024210	Sb09g024210	8.69	8.66	9.53	9.87	9.35
Ferulate 5-hydroxylase (F5H)						
F5H	Sb01g017270	6.83	7.06	6.88	6.64	6.57
F5H	Sb02g002630	6.10	6.15	6.53	6.78	6.04
F5H	Sb05g007210	5.34	5.30	6.93	6.35	5.73
Trans-cinnamate 4-hydroxylase (C4H)						
C4H	Sb02g010910	10.90	10.93	11.86	11.52	9.87
C4H	Sb03g038160	8.30	8.00	8.27	7.77	6.66
C4H	Sb04g017460	6.13	5.86	7.11	6.17	6.32
4-coumarate CoA ligase (4CL)						
4CL	Sb04g005210	10.21	10.17	11.06	10.64	9.07

Table A.9 (cont.) Expression of select phenylpropanoid-monolignol biosynthesis pathway genes. Please note that this is an excerpt of Table A9. The full table can be obtained at http://www.biomedcentral.com/1471-2229/14/35

4-coumarate CoA ligase (4CL)	Probes	seedling_shoot _PI455230_1	seedling_shoot _PI455230_2	seedling_root _PI455230	leaf_38DAP _ PI455230	shoot_tip_38DAP _ PI455230
4CL	Sb04g031010	7.83	7.61	7.12	8.50	7.41
4CL	Sb07g007810	3.29	5.36	6.88	5.31	3.17
4CL	Sb07g022040	8.69	8.87	9.33	10.45	7.53
4CL	Sb10g026130	10.34	10.48	11.20	10.71	9.02
Cinnamoyl CoA reductase (CCR)						
CCR	Sb01g028030	6.91	6.51	5.37	7.17	4.39
CCR	Sb03g029100	6.74	6.97	6.41	7.22	6.71
CCR	Sb09g029490	7.35	7.61	7.55	7.65	6.89
Hydroxyl-cinnamoyl CoA:shikimate/quinate-Hydroxyl-cinnamoyltransferase (HCT)						
HCT	Sb04g025760	10.78	10.85	11.63	10.15	9.32
Caffeoyl-CoA 3-O-methyltransferase (CCoAOMT)						
CCoAOMT	Sb02g027930	6.66	6.08	8.10	7.20	6.03
CCoAOMT	Sb07g028360	6.28	5.01	6.25	6.11	6.03
CCoAOMT	Sb07g028490	8.54	8.61	10.23	4.94	6.48
CCoAOMT	Sb07g028520	11.28	11.11	12.30	9.19	8.65
CCoAOMT	Sb07g028530	10.74	10.62	11.96	8.37	8.06
CCoAOMT	Sb10g004540	12.16	11.89	12.98	10.65	11.95
Caffeic acid O-methyltransferase (COMT)						
COMT	Sb07g003860	10.51	10.17	11.52	9.34	9.22
Cinnamyl alcohol dehydrogenase (CAD)						
CAD	Sb04g005950	6.99	7.38	8.52	7.22	6.56

* RMA-normalized log2 expression values

Table A.10 SAP Panel and population designation.

Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008
PI152651	milo/feterita	PI534155	durra	PI655993	kafir
PI17548	kafir	PI534157	zerazera/ caudatum	PI655994	kafir
PI24969	milo/feterita	PI534163	caudatum	PI655995	sudanense/ broomcorn
PI276837	NA	PI534167	NA	PI655996	milo/feterita
PI329435	zerazera/ caudatum	PI540816	milo/feterita	PI655997	milo/feterita
PI329440	kafir	PI542406	caudatum	PI655998	guinea/caudatum (W. Africa)
PI34911	caudatum	PI542718	durra	PI655999	sudanense/ broomcorn
PI35038	kafir	PI548797	caudatum	PI656000	zerazera/ caudatum
PI48770	kafir	PI552856	zerazera/ caudatum	PI656001	zerazera/ caudatum
PI533750	durra	PI552861	kafir	PI656002	kafir
PI533752	caudatum	PI561071	milo/feterita	PI656003	kafir
PI533754	guinea/bicolor	PI561072	zerazera/ caudatum	PI656004	guinea/caudatum (W. Africa)
PI533755	milo/feterita	PI561073	kafir	PI656005	NA
PI533757	caudatum	PI561472	zerazera/ caudatum	PI656006	caudatum
PI533758	zerazera/ caudatum	PI564163	zerazera/ caudatum	PI656007	milo/feterita
PI533759	caudatum	PI564164	zerazera/ caudatum	PI656008	zerazera/ caudatum
PI533760	zerazera/ caudatum	PI564165	zerazera/ caudatum	PI656009	zerazera/ caudatum
PI533761	caudatum	PI565121	zerazera/ caudatum	PI656010	guinea/caudatum (W. Africa)
PI533762	guinea/bicolor	PI566819	kafir	PI656011	NA
PI533766	guinea/caudatum (W. Africa)	PI574455	NA	PI656012	milo/feterita
PI533769	caudatum	PI576130	zerazera/ caudatum	PI656013	durra
PI533776	NA	PI576332	kafir	PI656014	sudanense/broom corn
PI533785	guinea/caudatum (W. Africa)	PI576333	kafir	PI656015	milo/feterita
PI533788	milo/feterita	PI576337	NA	PI656016	kafir
PI533789	caudatum	PI576339	kafir	PI656017	milo/feterita
PI533792	zerazera/ caudatum	PI576340	kafir	PI656018	kafir
PI533794	zerazera/ caudatum	PI576345	kafir	PI656019	kafir

Table A.10 (cont.) SAP Panel and population designation.

Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008
PI533799	NA	PI576347	sudanense/ broomcorn	PI656020	NA
PI533800	zerazera/ caudatum	PI576348	sudanense/ broomcorn	PI656021	milo/feterita
PI533807	caudatum	PI576349	sudanense/ broomcorn	PI656022	kafir
PI533810	durra	PI576350	NA	PI656023	kafir
PI533814	durra	PI576352	kafir	PI656024	zerazera/ caudatum
PI533821	guinea/caudatum (E. Africa/India)	PI576359	durra	PI656025	durra
PI533822	guinea/caudatum (E. Africa/India)	PI576364	kafir	PI656026	zerazera/ caudatum
PI533824	NA	PI576366	guinea/caudatum (E. Africa/India)	PI656027	caudatum
PI533830	NA	PI576373	milo/feterita	PI656028	kafir
PI533831	caudatum	PI576375	durra	PI656029	durra
PI533833	caudatum	PI576376	durra	PI656030	milo/feterita
PI533838	NA	PI576380	zerazera/ caudatum	PI656031	caudatum
PI533839	guinea/bicolor	PI576381	durra	PI656032	guinea/caudatum (E. Africa/India)
PI533841	NA	PI576385	kafir	PI656033	milo/feterita
PI533842	durra	PI576386	caudatum	PI656034	zerazera/ caudatum
PI533843	NA	PI576387	zerazera/ caudatum	PI656035	caudatum
PI533845	NA	PI576390	durra	PI656036	zerazera/ caudatum
PI533852	NA	PI576391	durra	PI656037	zerazera/ caudatum
PI533855	guinea/caudatum (E. Africa/India)	PI576393	guinea/caudatum (E. Africa/India)	PI656038	kafir
PI533856	durra	PI576394	kafir	PI656039	zerazera/ caudatum
PI533863	caudatum	PI576396	NA	PI656040	kafir
PI533866	guinea/caudatum (W. Africa)	PI576399	guinea/caudatum (E. Africa/India)	PI656041	milo/feterita
PI533869	guinea/caudatum (E. Africa/India)	PI576401	durra	PI656042	kafir
PI533871	guinea/caudatum (W. Africa)	PI576418	guinea/caudatum (W. Africa)	PI656043	zerazera/ caudatum
PI533876	caudatum	PI576422	kafir	PI656044	zerazera/ caudatum
PI533877	guinea/caudatum (W. Africa)	PI576425	durra	PI656046	durra
PI533878	NA	PI576426	durra	PI656047	durra

Table A.10 (cont.) SAP Panel and population designation.

Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008
PI533882	guinea/caudatum (W. Africa)	PI576428	zerazera/ caudatum	PI656048	zerazera/ caudatum
PI533901	caudatum	PI576434	NA	PI656049	kafir
PI533902	guinea/bicolor	PI576435	caudatum	PI656050	caudatum
PI533903	NA	PI576437	caudatum	PI656051	zerazera/ caudatum
PI533910	guinea/caudatum (E. Africa/India)	PI585291	milo/feterita	PI656052	kafir
PI533911	caudatum	PI585295	caudatum	PI656053	kafir
PI533912	guinea/caudatum (E. Africa/India)	PI595699	durra	PI656054	guinea/caudatum (W. Africa)
PI533913	caudatum	PI595702	kafir	PI656055	zerazera/ caudatum
PI533915	NA	PI595714	caudatum	PI656056	zerazera/ caudatum
PI533919	guinea/bicolor	PI595718	durra	PI656057	caudatum
PI533921	NA	PI595720	durra	PI656058	kafir
PI533924	NA	PI595739	zerazera/ caudatum	PI656059	zerazera/ caudatum
PI533927	guinea/bicolor	PI595740	milo/feterita	PI656060	kafir
PI533936	kafir	PI595741	guinea/caudatum (W. Africa)	PI656061	kafir
PI533937	kafir	PI595743	kafir	PI656062	NA
PI533938	guinea/caudatum (E. Africa/India)	PI595744	guinea/bicolor	PI656063	NA
PI533939	guinea/caudatum (E. Africa/India)	PI595745	caudatum	PI656064	NA
PI533940	kafir	PI597945	guinea/bicolor	PI656065	NA
PI533943	durra	PI597946	caudatum	PI656066	NA
PI533948	kafir	PI597949	guinea/caudatum (E. Africa/India)	PI656067	NA
PI533949	NA	PI597950	guinea/caudatum (E. Africa/India)	PI656068	NA
PI533954	NA	PI597951	guinea/caudatum (W. Africa)	PI656069	NA
PI533955	kafir	PI597952	caudatum	PI656070	NA
PI533956	guinea/caudatum (E. Africa/India)	PI597957	durra	PI656071	zerazera/ caudatum
PI533957	guinea/caudatum (E. Africa/India)	PI597958	kafir	PI656072	durra
PI533961	caudatum	PI597960	guinea/caudatum (W. Africa)	PI656073	milo/feterita
PI533962	caudatum	PI597961	zerazera/ caudatum	PI656074	guinea/bicolor
PI533964	zerazera/ caudatum	PI597964	zerazera/ caudatum	PI656075	zerazera/ caudatum

Table A.10 (cont.) SAP Panel and population designation.

Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008
PI533965	guinea/caudatum (E. Africa/India)	PI597965	NA	PI656076	zerazera/ caudatum
PI533967	caudatum	PI597966	caudatum	PI656077	durra
PI533970	caudatum	PI597967	zerazera/ caudatum	PI656078	guinea/bicolor
PI533972	caudatum	PI597968	guinea/caudatum (E. Africa/India)	PI656079	guinea/caudatum (W. Africa)
PI533976	caudatum	PI597971	caudatum	PI656080	guinea/caudatum (E. Africa/India)
PI533979	kafir	PI597972	NA	PI656081	guinea/bicolor
PI533980	caudatum	PI597973	guinea/bicolor	PI656082	guinea/bicolor
PI533985	zerazera/ caudatum	PI597976	guinea/caudatum (W. Africa)	PI656083	caudatum
PI533986	guinea/caudatum (E. Africa/India)	PI597980	caudatum	PI656085	guinea/bicolor
PI533987	caudatum	PI597982	caudatum	PI656086	guinea/bicolor
PI533989	caudatum	PI598069	sudanense/ broomcorn	PI656087	caudatum
PI533991	zerazera/ caudatum	PI598070	sudanense/ broomcorn	PI656088	durra
PI533996	caudatum	PI601816	zerazera/ caudatum	PI656089	durra
PI533997	guinea/caudatum (E. Africa/India)	PI607931	kafir	PI656090	guinea/caudatum (E. Africa/India)
PI533998	kafir	PI609456	NA	PI656091	durra
PI534009	durra	PI613536	caudatum	PI656092	durra
PI534021	durra	PI629034	milo/feterita	PI656093	guinea/caudatum (W. Africa)
PI534028	durra	PI629040	milo/feterita	PI656094	guinea/caudatum (W. Africa)
PI534037	guinea/bicolor	PI629059	kafir	PI656095	guinea/caudatum (W. Africa)
PI534046	NA	PI641824	kafir	PI656096	guinea/caudatum (W. Africa)
PI534047	NA	PI641836	kafir	PI656097	durra
PI534053	caudatum	PI641849	kafir	PI656098	caudatum
PI534054	kafir	PI641874	NA	PI656099	durra
PI534063	NA	PI642791	NA	PI656100	durra
PI534070	guinea/caudatum (W. Africa)	PI642793	NA	PI656101	milo/feterita
PI534075	guinea/caudatum (W. Africa)	PI642992	caudatum	PI656102	guinea/caudatum (E. Africa/India)
PI534079	caudatum	PI642998	sudanense/ broomcorn	PI656103	durra
PI534088	NA	PI651492	guinea/caudatum (E. Africa/India)	PI656104	durra

Table A.10 (cont.) SAP Panel and population designation.

Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008	Taxa	Subpopulation designation (Q=9) from Casa et al., 2008
PI534092	guinea/caudatum (W. Africa)	PI651496	caudatum	PI656105	caudatum
PI534096	guinea/bicolor	PI653616	caudatum	PI656106	caudatum
PI534097	kafir	PI653617	sudanense/ broomcorn	PI656107	zerazera/ caudatum
PI534099	caudatum	PI655970	kafir	PI656108	durra
PI534101	caudatum	PI655971	NA	PI656109	guinea/caudatum (E. Africa/India)
PI534104	caudatum	PI655972	kafir	PI656110	guinea/bicolor
PI534105	caudatum	PI655973	milo/feterita	PI656111	guinea/caudatum (E. Africa/India)
PI534108	caudatum	PI655974	kafir	PI656112	NA
PI534112	caudatum	PI655975	NA	PI656113	NA
PI534114	caudatum	PI655976	kafir	PI656114	NA
PI534115	zerazera/ caudatum	PI655977	zerazera/ caudatum	PI656115	NA
PI534116	guinea/bicolor	PI655978	guinea/caudatum (W. Africa)	PI656116	NA
PI534117	guinea/caudatum (E. Africa/India)	PI655979	milo/feterita	PI656117	NA
PI534123	guinea/bicolor	PI655980	sudanense/ broomcorn	PI656118	NA
PI534124	guinea/bicolor	PI655981	guinea/caudatum (W. Africa)	PI656119	NA
PI534127	guinea/bicolor	PI655982	NA	PI656120	NA
PI534128	durra	PI655983	kafir	PI656121	NA
PI534132	durra	PI655984	milo/feterita	PI659691	kafir
PI534133	durra	PI655985	milo/feterita	PI659692	zerazera/ caudatum
PI534135	durra	PI655986	milo/feterita	PI659693	zerazera/ caudatum
PI534137	caudatum	PI655987	kafir	PI659694	zerazera/ caudatum
PI534138	NA	PI655988	kafir	PI659695	zerazera/ caudatum
PI534139	sudanense/ broomcorn	PI655989	kafir	PI659696	zerazera/ caudatum
PI534144	guinea/caudatum (E. Africa/India)	PI655990	milo/feterita	PI659753	kafir
PI534145	guinea/caudatum (W. Africa)	PI655991	kafir		
PI534148	durra	PI655992	kafir		

Table A.11 Means and standard deviations of seed mineral concentrations from the Sorghum Association Panel (SAP) and Chromatin Association Panel (CHP) across three and one environments, respectively. Mineral concentration values are presented as mg kg^{-1}. Data represents an average of individual samples (n=234 and n=150) analyzed in tree (SAP) and one (CHP) separate experiments

Trait	SAP 2008 Lubbock, TX		SAP 2012 Puerto Vallarta, MX		SAP 2013, Field Rep #1 Florence, SC		CHP 2010, Puerto Rico	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
B	*10.4*	*2.33*	*13*	*6.28*	*14*	*6.48*	*12*	*2.41*
Na	*0.664*	*0.299*	*0.354*	*0.168*	*0.634*	*0.317*	*1.14*	*0.567*
Mg	*1700*	*199*	*1650*	*203*	*1490*	*250*	*1810*	*217*
Al	*0.633*	*0.322*	*0.156*	*0.0848*	*0.158*	*0.0618*	*0.828*	*0.566*
P	*3780*	*426*	*3750*	*479*	*2930*	*471*	*4040*	*457*
S	*963*	*101*	*915*	*97.4*	*841*	*131*	*1120*	*140*
K	*4380*	*659*	*4310*	*643*	*3350*	*591*	*4010*	*592*
Ca	*53.2*	*16.4*	*15.2*	*4.92*	*16.4*	*6.2*	*54.7*	*16.9*
Mn	*14.5*	*2.87*	*12.9*	*3.07*	*13.3*	*3.54*	*15.8*	*3.21*
Fe	*26.5*	*4.53*	*28.2*	*5.48*	*22.5*	*5.17*	*28.3*	*6*
Co	*0.0112*	*0.005*	*0.0065*	*0.0032*	*0.0034*	*0.0015*	*0.0119*	*0.00578*
Ni	*0.435*	*0.215*	*0.0818*	*0.0272*	*0.102*	*0.0532*	*0.495*	*0.242*
Cu	*3.43*	*1.13*	*1.42*	*0.582*	*2.65*	*1.05*	*5.71*	*1.92*
Zn	*25.7*	*5.6*	*19.5*	*3.96*	*17.3*	*4.27*	*21.3*	*5.61*
As	*0.0538*	*0.0159*	*0.0726*	*0.0278*	*0.096*	*0.0372*	*0.0537*	*0.0159*
Se	*1.33*	*0.273*	*1.21*	*0.375*	*1.37*	*0.545*	*1.3*	*0.673*
Rb	*1.42*	*0.229*	*0.892*	*0.482*	*2.38*	*0.68*	*2.28*	*0.631*
Sr	*1.48*	*0.475*	*0.381*	*0.13*	*0.215*	*0.0877*	*0.601*	*0.407*
Mo	*0.968*	*0.306*	*0.406*	*0.142*	*0.52*	*0.195*	*0.493*	*0.351*
Cd	*0.16*	*0.0915*	*0.0307*	*0.019*	*0.0404*	*0.0275*	*0.106*	*0.0672*

Table A.12 Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels

	B	Na	Mg	Al	P	S	K	Ca	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Cd
B	-																			
Na	-0.23																			
Mg	NS	NS																		
Al	-0.25	0.31	NS																	
P	0.25	NS	**0.71**	NS																
S	NS	NS	0.54	NS	**0.6**															
K	NS	NS	NS	NS	0.33	0.31														
Ca	NS	NS	NS	NS	0.32	0.23	0.29													
Mn	NS	NS	0.55	NS	0.56	0.43	NS	0.31												
Fe	NS	NS	0.54	NS	0.54	0.59	NS	NS	0.54											
Co	NS	NS	NS	NS	NS	0.23	NS	NS	0.23	0.27										
Ni	NS	NS	0.24	NS	NS	0.31	NS	NS	NS	0.34	0.51									
Cu	NS	NS	0.45	NS	0.51	0.4	NS	NS	0.39	0.52	NS	NS								
Zn	NS	NS	0.53	NS	0.54	0.43	NS	0.26	0.44	0.56	NS	0.3	0.52							
As	0.5	NS	0.26	NS	0.3	0.35	NS	NS	0.27	0.28	NS	NS	NS	0.22						
Se	0.26	NS	0.23	NS	0.22	0.24	NS	NS	NS	NS	NS	NS	NS	NS	0.28					
Rb	NS	NS	NS	NS	0.24	NS	0.42	NS	NS	NS	NS	NS	0.22	NS	NS	NS				
Sr	NS	NS	NS	NS	0.29	NS	0.26	**0.79**	0.25	NS	NS	NS	NS	0.22	NS	NS	NS			
Mo	NS	NS	NS	NS	0.22	0.32	NS	NS	NS	0.26	NS	NS	NS	0.26	NS	NS	NS	NS		
Cd	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-0.23	NS	-

SAP Rank Average Correlation coefficients

Table A.12 (cont.) Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels

	B	Na	Mg	Al	P	S	K	Ca	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Cd
B	-																			
Na	NS																			
Mg	NS	NS																		
Al	NS	NS	NS																	
P	NS	NS	**0.72**	NS																
S	NS	NS	**0.6**	NS	0.54															
K	NS	NS	NS	NS	0.41	0.4														
Ca	NS	NS	NS	NS	0.2	0.21	NS													
Mn	NS	NS	0.48	NS	0.5	0.47	NS	0.3												
Fe	NS	NS	0.48	NS	0.53	0.54	NS	NS	0.49											
Co	NS	NS	NS	0.22	NS	NS	NS	NS	NS	NS										
Ni	NS	NS	NS	0.21	NS	0.23	NS	NS	NS	0.22	**0.63**									
Cu	NS	NS	0.31	NS	0.37	0.36	NS	0.38	NS	0.42	NS	NS								
Zn	NS	NS	0.36	NS	0.33	0.36	NS	0.29	0.35	0.43	NS	NS	0.39							
As	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.2	0.21	NS	NS						
Se	0.2	NS	0.4	NS	0.4	0.43	NS	NS	0.32	0.27	NS	NS	NS	NS	NS					
Rb	0.24	NS	NS	NS	NS	0.24	0.44	NS	NS	NS	NS	NS	0.22	NS	NS	NS				
Sr	NS	NS	0.35	NS	0.35	0.24	NS	0.53	0.36	NS	NS	NS	NS	NS	NS	0.28	NS			
Mo	NS	NS	NS	NS	NS	0.29	NS	NS	NS	NS	0.23	NS	NS	NS	0.34	NS	NS	NS		
Cd	NS	0.21	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-

SAP 2008 Correlation coefficients

Table A.12 (cont.) Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels

	B	Na	Mg	Al	P	S	K	Ca	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Cd
B	-																			
Na	NS																			
Mg	NS	NS																		
Al	-0.43	NS	NS																	
P	0.31	NS	**0.76**	NS																
S	0.45	NS	**0.64**	NS	**0.73**															
K	NS	0.23	NS	NS	0.35	0.32														
Ca	NS	NS	NS	NS	0.25	0.25	0.27													
Mn	NS	NS	0.47	NS	0.48	0.37	NS	NS												
Fe	NS	NS	**0.6**	NS	0.55	0.54	NS	NS	0.35											
Co	NS	NS	NS	NS	NS	NS	NS	NS	0.33	NS										
Ni	NS	NS	NS	NS	NS	0.22	NS	NS	NS	0.32	0.36									
Cu	NS	NS	0.41	NS	0.47	NS	0.26	NS	0.22	0.52	NS	NS								
Zn	NS	NS	**0.62**	NS	0.59	0.55	NS	NS	0.37	**0.68**	NS	0.27	0.59							
As	0.91	NS	NS	-0.29	0.28	0.43	NS	NS	NS	NS	NS	NS	NS	NS						
Se	0.35	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.35					
Rb	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.22	NS	NS				
Sr	NS	NS	NS	NS	0.24	0.22	0.29	**0.88**	NS	NS	NS	NS	NS	NS	NS	NS	NS			
Mo	NS	NS	NS	NS	0.23	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS		
Cd	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-

SAP 2012 Correlation coefficients

Table A.12 (cont.) Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels

	B	Na	Mg	Al	P	S	K	Ca	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Cd
B	-																			
Na	NS																			
Mg	NS	NS																		
Al	NS	0.24	NS																	
P	0.23	NS	**0.8**	NS																
S	0.33	NS	**0.64**	NS	**0.7**															
K	NS	NS	NS	NS	0.42	0.37														
Ca	NS	NS	0.43	NS	0.38	0.44	0.22													
Mn	NS	NS	0.55	NS	0.55	0.49	NS	0.44												
Fe	NS	NS	0.57	NS	**0.6**	**0.69**	NS	0.3	**0.63**											
Co	-0.25	NS	NS	NS	NS	NS	NS	NS	0.34	0.31										
Ni	NS	NS	NS	NS	0.28	0.3	NS	NS	0.26	0.28	0.47									
Cu	NS	NS	0.49	NS	0.5	0.41	NS	0.21	0.41	0.54	NS	NS								
Zn	NS	NS	0.58	NS	0.53	0.55	NS	0.35	0.49	**0.62**	NS	0.25	**0.62**							
As	0.59	NS	NS	NS	0.32	0.38	NS	NS	0.25	0.29	NS	NS	NS	NS						
Se	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.2					
Rb	NS	NS	NS	NS	0.24	NS	0.4	NS	NS	NS	NS	NS	NS	NS	NS	NS				
Sr	NS	NS	0.35	NS	0.29	0.38	NS	**0.73**	0.4	NS	NS	NS	NS	0.31	NS	NS	NS			
Mo	NS	NS	NS	NS	0.22	0.37	NS	NS	NS	0.23	NS	NS	0.24	0.25	NS	NS	-0.23	NS		
Cd	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-0.21	NS	-

SAP 2013 Correlation coefficients

Table A.12 (cont.) Correlation coefficients among seed mineral concentrations using average data across replicates from SAP and CHP association panels

	B	Na	Mg	Al	P	S	K	Ca	Mn	Fe	Co	Ni	Cu	Zn	As	Se	Rb	Sr	Mo	Cd
B	-																			
Na	NS																			
Mg	NS	NS																		
Al	NS	NS	NS																	
P	NS	NS	**0.78**	NS																
S	NS	NS	0.51	NS	0.56															
K	NS	NS	NS	NS	NS	NS														
Ca	NS	NS	0.35	NS	NS	NS	NS													
Mn	NS	NS	0.54	NS	0.51	NS	NS	0.53												
Fe	NS	NS	0.33	NS	0.49	0.36	NS	NS	0.37											
Co	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS										
Ni	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.33									
Cu	NS	NS	NS	NS	0.36	**0.69**	NS	NS	NS	0.37	NS	NS								
Zn	NS	NS	0.41	NS	NS	NS	NS	0.35	NS	NS	NS	NS	0.37							
As	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS						
Se	NS	-0.33	NS	0.35	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.39	NS					
Rb	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-0.39				
Sr	NS	NS	NS	0.4	NS	NS	NS	0.43	NS	NS	NS	NS	NS	0.43	NS	**0.77**	-0.37			
Mo	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	0.35	NS	**0.73**	-0.33	**0.68**		
Cd	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-

CHP 2010 Correlation coefficients

Table A.13 Significant SNPs identified in the 2008 SAP experiment.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
S6_46236341	S3_4670311	S3_2005315	S1_7268318	S1_58901330	S2_8464347	S1_4192743	S1_42266436	S1_62913850
S6_46589170	S3_65723541	S6_47289998	S10_56869835	S1_59694228	S2_8756240	S10_57861319	S1_46430264	S1_62913851
S8_50935291	S6_33729035	S6_47378903	S10_59844644	S1_66757591	S6_950607	S2_4575286	S1_46430278	S1_67073677
S8_53705782	S6_33729050	S6_47394326	S2_61133102	S10_60256055	S9_5215697	S2_4575295	S1_46430282	S1_67073678
	S6_38823793		S2_67972537	S10_60256056	S9_57071090	S2_4630010	S6_56315903	S2_20233992
	S6_38836159		S2_67972541	S10_60256058	S9_57169734	S2_66788690	S8_49332942	S2_32913651
	S6_39631053		S2_68029839	S2_2593395	S9_57169736	S2_67686143	S9_50823409	S2_62906408
	S6_4854341		S2_68725882	S9_2072218	S9_57169738	S3_2815263		S2_62906428
	S6_4874260		S2_68746019		S9_57169739	S3_56117537		S2_62906429
	S6_4893258		S6_38495962		S9_57169740	S3_64702081		S2_63083669
	S6_5012832		S6_38495989		S9_57222599	S3_70141162		S3_64538301
	S6_5038109		S6_48883295		S9_57793559	S3_70141169		S3_64563673
	S6_5049251		S6_48883486		S9_57793560	S3_70151489		S3_64563679
	S6_5504814		S6_53244423		S9_57827956	S4_1663760		S3_70568050
	S6_5546431		S7_63641183		S9_58123978	S4_1665086		S4_14747062
	S6_5635644		S7_9855356		S9_58331590	S5_60567301		S4_20838277
	S6_5707094		S8_49147915		S9_58373621	S6_54572292		S4_23142621
	S6_6637896		S9_50261101		S9_58419483	S6_57560442		S4_2488750
	S6_6856224		S9_56681292		S9_58419495	S8_14305708		S4_2495331
	S7_58301681		S9_57940747		S9_58432544	S8_14446591		S6_58323365
	S7_58301683		S9_57958750		S9_58432571	S9_567209		S8_39744152
	S7_58390034		S9_58050897		S9_58445767	S9_58373628		S8_40251965
	S7_58390037				S9_58509245			S9_1452516
	S7_59050134				S9_58509247			
	S7_59361929				S9_58593789			
	S7_60031456				S9_58595468			
	S7_60031458				S9_58600625			

Table A.13 (cont.) Significant SNPs identified in the 2008 SAP experiments.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
	S7_60031459				S9_58690753			
	S7_60031509				S9_58713956			
					S9_58713967			
					S9_58713982			
					S9_58726757			
					S9_58726846			
					S9_58778466			
					S9_58790930			
					S9_58821537			
					S9_58889403			
					S9_58951376			
					S9_59030912			
					S9_59369385			
Mg25	**Mn55**	**Mo98**	**Na23**	**Ni60**	**P31**	**Rb85**	**S34**	**Se82**
S10_3500536	S3_53492481	S1_17179668	S2_3409802	S1_15918710	S1_14807427	S10_11669823	S1_43881380	S1_59429974
S10_45587987	S3_69129602	S1_19448647	S2_67011653	S4_10064032	S2_63082960	S10_11669832	S1_54221148	S1_59429981
S2_74431920	S4_4181353	S2_19504377	S2_67011658	S4_20111973	S2_63082961	S2_71792789	S10_3500536	S4_4659107
S5_49658462	S6_38836159	S2_19504389	S2_72752460	S4_20111987	S2_67935905	S3_61745847	S10_3775055	S6_20939153
S5_56964578		S2_76455695	S3_4451321	S4_20112010	S5_45245599	S4_2400578	S4_5931356	S6_20939187
S6_56315903		S3_64823106	S3_4451330	S5_53728030	S5_49638362	S4_2688589	S5_45245599	S8_54921913
S8_1035307		S4_48603048	S5_46320421	S6_53175238	S5_49638381	S4_2688595	S5_53704903	S8_54921914
S8_52240073		S4_60104546	S8_2367673	S7_63062717	S5_49658462	S6_50357871	S5_56205069	
		S6_39631053	S8_2432745	S9_11977404		S9_47318662	S5_56410206	
		S6_42736415	S8_2432752				S8_2347100	
		S6_6248047	S8_47302555				S9_56635891	
		S9_2668599	S8_47302567				S9_56670555	
							S9_56670561	
							S9_56712191	

Table A.13 (cont.) Significant SNPs identified in the 2008 SAP experiments.

Sr88	Zn66
S1_1223616	S1_1282948
S1_1223620	S1_65255626
S10_14699742	S2_18030873
S10_371094	S5_52461982
S10_55295019	
S10_7425578	
S2_68718931	
S2_70558866	
S3_73387638	
S3_73536835	
S3_73539758	
S3_73662300	
S4_5232791	
S4_9467568	
S6_1984859	
S6_44860205	
S7_4921152	
S7_543144	
S7_62098472	
S8_54766973	
S8_54766975	
S8_922477	
S8_930712	
S9_56681292	

Table A.14 Significant SNPs identified in the 2012 SAP experiments.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
S6_2171997	S10_1870292	S10_1870255	S1_13117127	S2_8851899	S1_21752163	S1_66237143	S1_19766414	S1_4711330
S6_2172003	S10_55262277	S10_1870292	S1_13117140	S2_8883293	S5_42237849	S2_52199692	S1_19766422	S1_4897115
S6_2172141	S2_8478661	S3_66894535	S1_20327511	S6_42008432	S5_42269511	S4_35905724	S1_19766440	S1_5566882
S6_3418377	S2_8478699	S4_51664963	S1_21252486		S5_42296486	S4_36974160	S2_61935877	S1_5566884
	S2_8478911	S6_2171997	S1_23942439		S5_42296492	S4_37362071	S4_6412732	S1_5566886
	S2_8478918	S6_2172003	S1_6887826		S5_42508220	S4_37572439	S4_6426225	S10_15710072
	S6_2171997		S1_6891752		S7_24605131	S4_37598785	S5_51030574	S3_11671601
	S6_2172003		S1_6957656		S8_1939064	S4_38723197	S5_51030587	S3_11671602
			S1_7268318		S8_1939066	S4_39679477	S6_254831	
			S1_7355484			S4_40473989	S9_2457721	
			S1_7355554			S4_41150300		
			S1_7355556			S6_59788013		
			S1_7385618					
			S1_7385621					
			S1_7409871					
			S1_7416761					
			S1_7416788					
			S1_7454989					
			S1_7479328					
			S1_7479338					
			S2_52598094					
			S2_52598095					
			S2_59799911					
			S2_64668865					
			S2_64668867					
			S2_64668870					
			S2_64668871					

Table A.14 (cont.) Significant SNPs identified in the 2012 SAP experiments.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
			S3_69387872					
			S4_52526364					
			S5_3243616					
			S5_5664675					
			S8_3273535					
			S8_37877207					
Mg25	**Mn55**	**Mo98**	**Na23**	**Ni60**	**P31**	**Rb85**	**S34**	**Se82**
S1_25585419	S1_26914874	S1_23934576	S1_24354821	S1_18898717	S1_64935466	S10_1782172	S10_4130467	S10_15672012
S1_64935466	S1_26985947	S10_57534008	S1_46799306		S1_64935513	S2_2647310	S5_42545335	S10_16378173
S10_47832699	S1_27019468	S3_15723150	S1_56960508		S10_8205070	S2_2647315	S8_2982477	S10_16378176
S10_47832700	S1_27129528	S3_64874580	S10_56828763		S2_64294886	S2_5051111	S8_3994475	S2_6143436
S10_55413223	S1_68067401	S3_64874613	S10_56828766		S3_69177346	S2_61828141		S2_62858510
S10_6944717	S10_56596658	S3_64874614	S2_6542470		S5_41756651	S6_52458936		S2_6588539
S2_14699473	S2_1275051	S3_72700087	S2_72421123			S8_51850803		S2_6588541
S2_72899590	S3_4338380	S3_73474329	S2_72421167			S8_6186108		S3_56997210
S2_72899620	S6_44514232	S4_35485734	S3_1776750			S9_1626685		
S3_14245882	S8_11922791	S4_55414091	S3_1776751					
S3_14245897		S4_55414097	S3_65451666					
S3_14245954		S5_6173635	S5_12284079					
S3_14245999		S6_59739227	S7_54739020					
S3_64239050		S6_59739236	S8_542052					
S4_5679211		S6_59739254						
S4_6426225		S6_59739279						
S4_7244474		S7_11143268						
S4_7244483		S7_11403377						
S6_1278954		S8_13350137						
S6_1279029		S8_13350157						
S6_1279030		S8_48608273						

Table A.14 (cont.) Significant SNPs identified in the 2012 SAP experiments.

Mg25	Mn55	Mo98	Na23	Ni60	P31	Rb85	S34	Se82
S6_1279097		S9_58432526						
S6_1279163								
S6_1283219								
S7_3702075								
S8_38559035								
S8_38559057								

Sr88	Zn66
S1_6957656	S1_11809098
S1_7268318	S1_19767173
S2_64668865	S1_25585419
S2_64668867	S1_55530866
S2_64668870	S1_5865548
S2_64668871	S10_4433796
S3_69387872	S10_4433829
S3_70980898	S10_4439772
S3_70980899	S10_5456890
S3_70980900	S2_72899590
S3_73089352	S2_72899620
S4_51717721	S4_6818081
S4_52582137	S9_17928193
S5_3243616	S9_39419630
S5_54611172	
S7_57390315	
S7_57390498	
S7_57394182	
S8_43045057	
S8_43045060	

Table A.15 Significant SNPs identified in the 2013 SAP experiments.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
S2_2593395	S1_46880052	S1_10939262	S1_3515686	S2_8492870	S1_68434340	S2_65569563	S1_22989645	S4_55477250
S2_2605594	S1_46880058	S1_57900684	S1_3515687	S2_8883293	S6_1273519	S8_53510880	S1_22989647	S4_55481016
S2_3406009		S3_62655346	S1_57677209	S2_8917154	S6_5554189		S1_42930927	S4_55481026
		S7_4239563	S1_6658763	S7_10013541	S6_5599641		S1_7268286	S4_56798220
		S7_4239571	S1_6887826				S1_7268324	S9_3205116
		S7_62446449	S1_6891752				S1_7268325	S9_50341821
		S9_7549209	S1_6957656				S10_46952041	S9_55342131
			S1_7051267				S2_6039247	S9_863539
			S1_7268318				S3_9797814	S9_863543
			S1_7320214				S3_9797968	
			S1_7321546				S4_5931356	
			S1_7321553				S6_51031988	
			S1_7355554				S6_51077165	
			S1_7355556				S6_58736895	
			S1_7356148				S6_58737019	
			S1_7455160				S9_51024449	
			S1_8064876					
			S2_20146970					
			S2_9842808					
			S3_52296657					
			S4_4664585					
			S9_43995476					
Mg25	**Mn55**	**Mo98**	**Na23**	**Ni60**	**P31**	**Rb85**	**S34**	**Se82**
S1_50576720	S1_52681005	S1_26205140	S1_51505976	S1_12950685	S1_50577583	S1_43750629	S1_47675382	S10_12633125
S1_50577567	S1_66324608	S1_26205142	S1_51506437	S10_10589487	S1_50577591	S3_12614651	S10_45922516	S3_3955251

Table A.15 (cont.) Significant SNPs identified in the 2013 SAP experiments.

Mg25	Mn55	Mo98	Na23	Ni60	P31	Rb85	S34	Se82
S1_50577583	S10_40016943	S3_57626685	S10_16061375	S10_11025513	S10_48726890		S10_45926714	S5_53978073
S1_50577591	S10_42265325	S3_57626706	S10_16061377	S2_75101499	S2_7973144		S10_45926740	S8_47750019
S1_64935466	S10_42455170	S3_57626712	S10_18197644	S4_59989542	S5_9003500		S10_45926741	S9_50745150
S1_64935513	S10_42478803	S3_57626724	S10_7717763		S8_41763641		S10_46754053	
S10_48726890	S10_42545335	S3_57627814	S10_7732382		S8_41763665		S10_46754092	
S2_56541850	S10_42915193	S4_12452286	S3_4755673		S8_41763674		S3_9797968	
S2_716542	S10_42985179	S4_12452312	S5_13477302					
S6_49754782	S3_6112767	S4_12588668	S5_13533594					
S6_55096391	S4_10547505	S4_44992538	S5_13534544					
S7_55192577	S6_1558769	S9_17080763	S5_51147208					
S7_55242120		S9_17637468	S5_51147211					
		S9_17688137	S5_51147244					
		S9_18133268	S5_51147249					
		S9_18423658	S5_5191343					
		S9_18706368	S6_61146826					
		S9_18771206	S6_61146855					
		S9_19134113						
		S9_20543306						
		S9_20703240						
		S9_21084139						
		S9_21085301						
		S9_22230425						
		S9_22230426						
		S9_22578810						
		S9_22723887						
		S9_24615137						
		S9_25108706						
		S9_25506152						

Table A.15 (cont.) Significant SNPs identified in the 2013 SAP experiments.

Mg25	Mn55	Mo98	Na23	Ni60	P31	Rb85	S34	Se82
		S9_25930052						
		S9_26643357						
		S9_27052179						
		S9_30764631						
		S9_31012869						
		S9_31540947						
		S9_31928959						
		S9_32843236						
		S9_33531645						
		S9_34408285						
		S9_35308889						
		S9_35770541						
		S9_35873516						
		S9_36252593						
		S9_38236344						
		S9_38568059						
		S9_39213829						
		S9_40789769						
		S9_43029100						

Sr88	Zn66	Zn66 (cont.)
S1_72480496	S1_11421511	S9_1851297
S10_53809968	S1_42222656	S9_4011156
S3_52296657	S10_43196994	
S3_5493754	S10_46252473	
S3_56769683	S2_55856498	
S3_70115846	S2_56017352	
S6_5805174	S4_55761052	
S9_6537326	S4_968545	

Table A.16 Significant SNPs identified in the SAP experiments based on rank average. Lines were ranked by mineral concentration, and GWAS was carried out on the average rank across 4 SAP locations

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
S10_3385499	S2_884243	S3_62655346	S1_51127889	S1_52575492	S2_73416536	S10_51868304	S1_19766414	S2_55832959
S10_3385505	S3_20086399	S4_52068874	S1_51127898	S10_56217144	S2_8464347	S10_51919930	S10_11100382	S4_21416067
S10_3385509	S3_20086471		S1_57677209	S10_56217186	S2_8738381	S2_66788690	S10_11117141	S6_2119293
S2_17813725	S3_20086540		S1_7268318	S10_56217255		S2_67686143	S10_1134998	S6_45464405
S2_17814453	S4_47752145		S2_61435085	S10_56236951		S6_59788013	S10_45926714	S6_45965598
S2_17830361	S4_51360594		S2_62999349	S2_13097435		S8_2105096	S10_45926740	S6_45971634
S2_17996567	S4_52052545		S2_63085287	S2_13097464		S8_2105101	S10_45926741	S6_46047456
S2_18214050			S2_63085461	S2_8883293		S8_49287477	S3_12122609	S6_46047502
S2_18520773			S2_63085503	S3_64938319			S3_71815187	S6_46049360
S2_18586646			S2_65758983	S5_10119658			S6_59652578	S6_46049361
S2_18796715			S2_65833255	S5_9588341			S9_8071635	S6_46053843
S2_18839386			S2_65899225	S5_9588371				S6_46063500
S2_19268078			S2_65899226	S5_9588375				S6_46063503
S2_19342789			S2_65997118	S8_44734488				S6_46065775
S2_19651840			S2_66184472	S9_58345051				S6_46065776
S2_19880265			S2_67431062	S9_58545487				S6_46065777
S2_19939195			S2_67972541					S6_46137219
S2_20718883			S2_68718931					S6_48568089
S2_21005895			S2_69066860					S6_48568133
S2_76564899			S3_48856543					S6_58328674
S3_4512457			S3_48871983					S6_58508824
S7_10028442			S3_4969531					S6_58521354
			S3_8468708					S9_10939684
			S3_8473833					
			S3_8500715					

Table A.16 (cont.) Significant SNPs identified in the SAP experiments based on rank average. Lines were ranked by mineral concentration, and GWAS was carried out on the average rank across 4 SAP locations

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
			S3_8518181					
			S3_8653912					
			S4_52526364					
			S4_53331810					
			S5_2607230					
			S5_4403554					
			S6_48883486					
			S9_50261101					
Mg25	**Mn55**	**Mo98**	**Na23**	**Ni60**	**P31**	**Rb85**	**S34**	**Se82**
S1_64935466	S1_18898717	S10_52670060	S1_2263815	S1_15038536	S1_51089486	S1_56343767	S1_39790822	S1_10901761
S2_1496872	S1_18937612	S10_54917226	S1_2263842	S1_15038537	S1_64935466	S1_56343768	S1_71251842	S2_60542958
S2_56541850	S1_67656597	S2_7127502	S5_18831660	S3_10110935	S1_64935513	S6_51614195	S10_1134998	S2_60570113
S2_716542	S10_42478803	S3_64823106	S5_18953099	S4_59533264	S1_66747492	S6_51614226	S3_70638233	S2_69840521
S4_38723197	S3_4290756	S4_16304537	S5_48918062	S4_60949342	S2_63082960	S6_52458936		S2_70103895
	S3_66960028	S4_17837791	S5_49638889	S4_61023448	S2_63082961	S8_6186108		S2_70105741
	S4_4181353	S4_19711296	S5_51806026	S4_61060973	S2_8359329	S9_42990123		S2_70310459
	S5_4230376	S4_19972269	S5_51806047	S4_61424502	S3_3957811	S9_42990156		S8_4254335
	S5_53895485	S4_35485734	S7_5953365	S4_6227570				S8_54921913
	S5_737632	S4_45937677		S4_63762993				S8_54921914
	S5_737646	S4_50093056		S4_63944216				S8_5610037
		S4_50093058		S4_63944217				S9_58221105
		S4_50418554		S5_1897447				
		S4_55037771		S6_53175238				
		S6_43237707		S9_22041141				
		S6_43237728		S9_24604814				
		S9_56708944						

Table A.16 (cont.) Significant SNPs identified in the SAP experiments based on rank average. Lines were ranked by mineral concentration, and GWAS was carried out on the average rank across 4 SAP locations

Sr88	Zn66
S1_7268318	S1_11421511
S2_62776928	S1_3221833
S2_68718931	S1_55530866
S2_69066860	S1_5865548
S2_69072967	S1_8000153
S2_69136016	S1_8002737
S2_69141817	S1_8002755
S2_69718256	S1_8002757
S3_56769683	S1_8053602
S3_58850829	S10_2178300
S3_69160562	S10_2300100
S3_71693122	S10_4433796
S3_73237503	S10_4433829
S4_5986109	S10_4435299
S4_5986126	S2_74762860
S5_2607230	S3_48874773
S5_3243616	S3_48875359
S8_44579921	S3_55682865
	S3_8694301
	S4_15028075
	S7_6880986
	S9_1851297

Table A.17 Significant SNPs identified in the 2010 CHP experiments.

Al27	As75	B11	Ca43	Cd111	Co59	Cu65	Fe57	K39
S1_6962011	S1_72473385	S1_63758193	S10_1617546	S2_56964725	S1_18765975	S1_15192612	S1_14630964	S1_14683435
S1_6962695		S10_38278163	S10_1617577	S2_56964734	S1_70722797	S1_25938382	S1_61175358	S10_55144781
S1_6962725		S2_64527419	S2_71458490	S2_56965657	S1_71607670	S1_53676057	S1_66980315	S2_1462274
S2_62511099			S4_51026793	S3_15078786	S10_59419647	S2_77217134	S10_5117117	S2_1462287
S3_66947352			S4_51026815	S3_49632244	S10_59435356	S3_57158997	S5_5229463	S3_8288203
S4_159132			S6_42350717	S3_50067993	S2_70618460	S4_58894975	S6_54910895	S5_59463541
S4_66562420			S8_49961265	S3_50399038	S3_72710672	S4_61835734	S6_56043995	S6_6756839
S6_53175238			S9_46183390	S3_7307520	S3_9992147	S5_1449327		S7_63763682
S6_55835821				S4_60638734	S5_57133211	S6_48526818		
S7_4859311				S5_3578238	S6_1091969	S6_61788062		
S7_55741000					S9_57275850	S6_61819228		
S7_62511464						S6_61845947		
S7_62511932								
S7_62513213								
S9_11976880								
S9_11978287								
S9_15312151								
S9_45099279								
S9_45116164								
S9_45116513								
S9_56410262								
S9_9650663								
Mg25	**Mn55**	**Mo98**	**Na23**	**Ni60**	**P31**	**Rb85**	**S34**	**Se82**
S1_46724315	S10_1149373	S1_62031867	S10_52569331	S1_12682892	S1_46521811	S10_4598620	S10_11669823	S1_1872900
S1_60064889	S10_1150238	S10_55302130	S4_7343272	S2_69718256	S1_57278235	S2_70112500	S10_11669832	S3_20038881
S2_61811047	S2_69841753	S4_4759604	S5_1466773	S3_4521544	S1_59527863	S2_70112560	S10_11669934	S3_5761928
S2_7356180	S2_75877655	S5_59770572		S3_48788390	S3_59223900	S2_76968828	S10_53042685	S3_73485171

Table A.17 (cont.) Significant SNPs identified in the 2010 CHP experiments.

Mg25	Mn55	Mo98	Na23	Ni60	P31	Rb85	S34	Se82
S4_53331324	S3_20029499	S6_54473620		S3_63571823	S3_64454058	S2_76991378	S10_53042706	S4_35783501
S5_62182157	S4_61564973	S8_45864477		S3_69580892	S3_64466194	S3_4347506	S2_61059179	S4_54382918
S7_22827989	S4_61564974			S3_70567940	S4_54145966	S3_56907419	S2_77217134	S4_55567767
S7_27800114	S5_60278713			S4_5671046	S4_63313730	S3_56907431	S4_55003477	S5_2291591
S7_3708522	S5_60278717			S7_62836516	S6_1340965	S4_1390604	S4_8308692	S5_56024096
S7_9663126	S6_42214158			S8_2261578	S6_53452505	S4_5253475	S6_42640203	S5_56024334
	S6_47249244				S6_53784270	S4_711577	S6_48526818	S6_2498495
	S6_50089079					S4_711579	S8_445813	S6_2498536
	S7_2101527					S4_711606		S6_2498539
	S7_2330182					S5_56307499		S7_57712049
	S8_52905746					S8_5552498		S8_46002703
	S8_52905782							S8_46002706
	S8_52905789							S8_54145206
								S9_2696814
								S9_2696820

Sr88	Zn66
S1_1778933	S2_72243248
S3_71789162	S3_69177346
S9_14983325	S5_58246488
S9_48012543	S6_58251195
S9_48012573	S6_58251198
S9_53006284	S6_58251208
	S8_2447301

Table A.18 Shared significant SNPs across SAP experiments

Mineral	SAP 2008	SAP 2012	SAP 2013	SAP Ranked Average
B11	-	-	S3_62655346	S3_62655346
Na23	-	-	-	-
Mg25	-	**S1_64935466**	**S1_64935466**	**S1_64935466**
	-	-	S2_56541850	S2_56541850
	-	-	S2_716542	S2_716542
Al27	-	-	-	-
P31	-	S1_64935466	-	S1_64935466
	-	S1_64935513	-	S1_64935513
	S2_63082960	-	-	S2_63082960
	S2_63082961	-	-	S2_63082961
S34	-	-	-	-
K39	-	-	-	-
Ca43	-		S1_57677209	S1_57677209
	-	S1_6887826	S1_6887826	-
	-	S1_6891752	S1_6891752	-
	-	S1_6957656	S1_6957656	-
	S1_7268318	**S1_7268318**	**S1_7268318**	**S1_7268318**
	S2_67972541	-	-	S2_67972541
	-	S1_7355554	S1_7355554	-
	-	S1_7355556	S1_7355556	-
	-	S4_52526364	-	S4_52526364
	S6_48883486	-	-	S6_48883486
	S9_50261101	-	-	S9_50261101
Mn55	-	-	S10_42478803	S10_42478803
	S4_4181353	-	-	S4_4181353
Fe57	-	S1_19766414		S1_19766414
Co59	S2_8464347	-	-	S2_8464347

Table A.18 (cont.) Shared significant SNPs across SAP experiments

Mineral	SAP 2008	SAP 2012	SAP 2013	SAP Ranked Average
Ni60	S6_53175238	-	-	S6_53175238
Cu65	S2_66788690	-	-	S2_66788690
	S2_67686143	-	-	S2_67686143
	-	S6_59788013	-	S6_59788013
Zn66	-	-	S1_11421511	S1_11421511
	-	S1_55530866	-	S1_55530866
	-	S1_5865548	-	S1_5865548
	-	S10_4433796	-	S10_4433796
	-	S10_4433829	-	S10_4433829
	-	-	S9_1851297	S9_1851297
As75	-	-	-	-
Se82	S8_54921913	-	-	S8_54921913
	S8_54921914	-	-	S8_54921914
Rb85	-	S6_52458936	-	S6_52458936
	-	S8_6186108	-	S8_6186108
Sr88	-	S1_7268318	-	S1_7268318
	S2_68718931	-	-	S2_68718931
	-	-	S3_56769683	S3_56769683
	-	S5_3243616	-	S5_3243616
Mo98	S3_64823106	-	-	S3_64823106
	-	S4_35485734	-	S4_35485734
Cd111	-	S2_8883293	S2_8883293	S2_8883293

Table A.19 Significant SNPs shared in multiple elements

SAP 2008- SNPs above or near the bonferroni (< 1e-4)	**Element1**	**Element2**	
S6_38836159	As	Mn	
S6_39631053	As	Mo	
S9_56681292	Ca	Sr	
S6_56315903	Fe	Mg	
S10_3500536	Mg	S	
S5_49658462	Mg	P	
S5_45245599	P	S	
SAP 2012- SNPs above or near the bonferroni (< 1e-4)	**Element1**	**Element1**	**Element3**
S6_2171997	Al	As	B
S6_2172003	Al	As	B
S10_1870292	B	As	
S1_6957656	Ca	Sr	
S1_7268318	Ca	Sr	
S2_64668865	Ca	Sr	
S2_64668867	Ca	Sr	
S2_64668870	Ca	Sr	
S2_64668871	Ca	Sr	
S3_69387872	Ca	Sr	
S5_3243616	Ca	Sr	
S4_6426225	Fe	Mg	
S1_25585419	Mg	Zn	
S1_64935466	Mg	P	
S2_72899590	Mg	Zn	
S2_72899620	Mg	Zn	

Table A.19 (cont.) Significant SNPs shared in multiple elements

SAP 2013- SNPs above or near the bonferroni (< 1e-4)	**Element1**	**Element2**	
S3_52296657	Ca	Sr	
S3_9797968	Fe	S	
S1_50577583	Mg	P	
S1_50577591	Mg	P	
S10_48726890	Mg	P	
Rank Average- SAP populations	**Element1**	**Element2**	
S1_7268318	Ca	Sr	
S2_68718931	Ca	Sr	
S2_69066860	Ca	Sr	
S5_2607230	Ca	Sr	
S10_1134998	Fe	S	
S1_64935466	Mg	P	
CHP 2010- SNPs above or near the bonferroni (< 1e-4)	**Element1**	**Element2**	
S2_77217134	Ca	Sr	
S6_48526818	Ca	Sr	

Table A.20 Sorghum forage, fuel and 'NREL' NIR trait curves

Forage Traits	Fuel Traits	NREL traits
ADF [% DM]	BTU Dry [BTU/lb]	Arabinan [%]
Crude Protein[%DM]	BTU MMF [BTU/lb]	Ash_NREL [%]
IVTDMD 30 [% DM]	BTU wet [BTU/lb]	EToH Extract [%]
K2O [mg/kg]	Carbon [%]	Galactan [%]
Lignin [% DM]	Cl [% wt]	Lignin_NREL [%]
ND-ICP [% DM]	Fixed Carbon [%]	NonStrucInorg [%]
NDF [% DM]	Fuel_Ash [%]	NonStruc_Protein
NDFD 30 [% DM]	Hydrogen [%]	Protein_NREL[%]
NFC [%]	K [mg/kg]	Struc_Protein [%]
Neg [Mcal/cwt]	K2O [%]	Struct Inorg [%]
Nel [Mcal/cwt]	Na [mg/kg]	Sucrose_NREL [%]
Nem [Mcal/cwt]	Na2O [%]	Water_Extracts [%]
S [% wt]	Nitrogen [%]	Whole_Starch [%]
TDN [%]	Oxygen_Diff [%]	cellulose [%]
ash_Forage [% wt]	S [%]	glucan [%]
	SiO2 [%]	soluble [%]
	Total halogens [%]	struct_starch [%]
	Volatile Matter [%]	structural [%]
		xylan [%}

APPENDIX B

SUPPLEMENTARY FIGURES

.

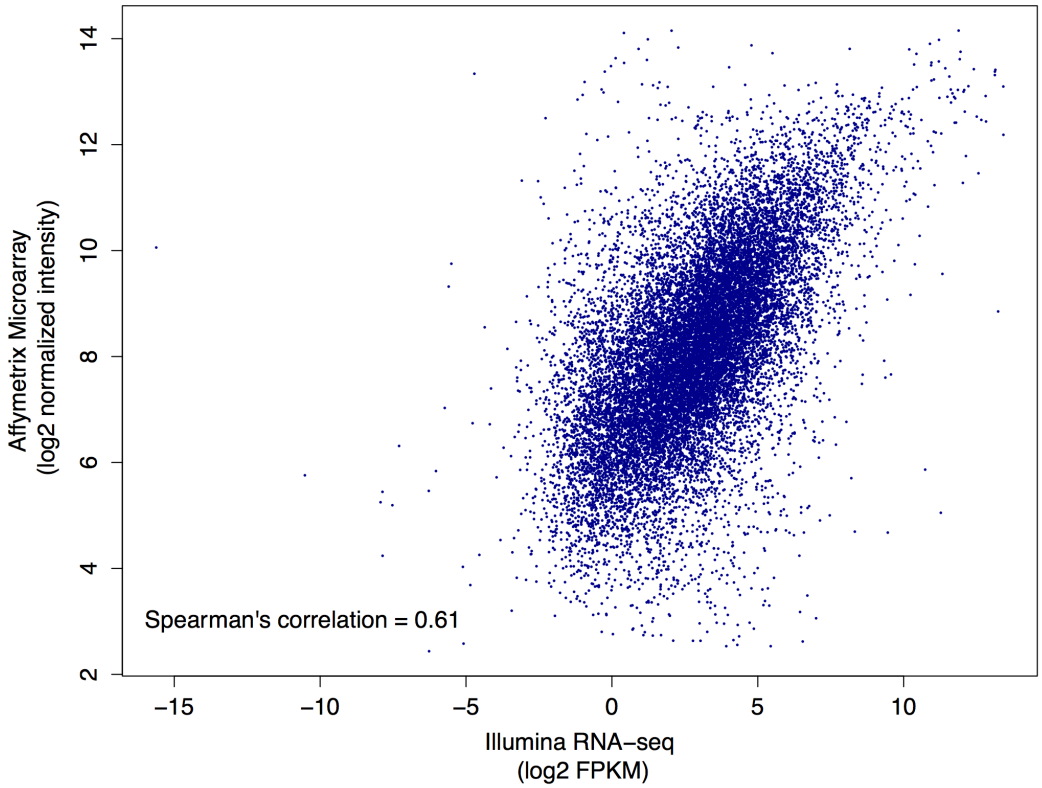


Figure B.1 Correlation of RNA expression between Illumina RNA sequencing and Affymetrix GeneChip microarray platform. * Each point represents a sorghum gene identified in grain sorghum leaf tissue of BTx623 by RNA-Sequencing, and by microarray in R159. RNA-Sequencing expression levels were measured using RPKM [19] and array levels were measured using the mean intensity of sense probes within exons. The Spearman's coefficient is 0.61, which is consistent with previous studies and indicates that the platforms correlate well on similar samples.

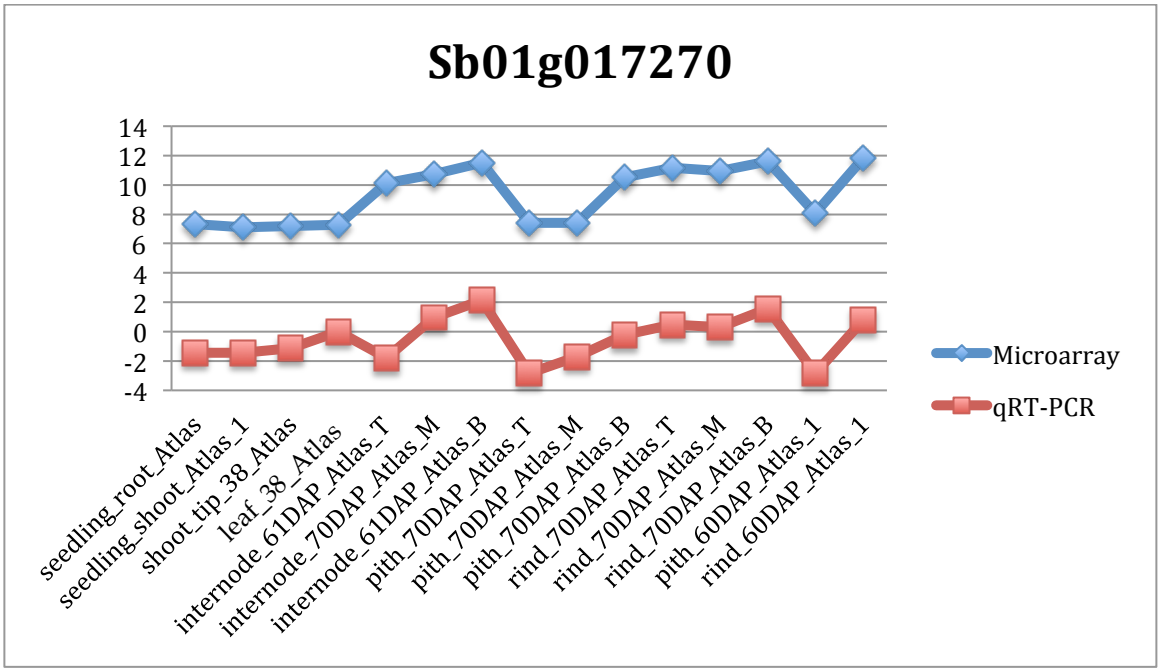


Figure B.2 Expression dynamics across multiple tissue types detected by microarray and qRT-PCR. * Expression levels of Sb01g01720 varied across tissue types. The vertical axis shows the $\log_2$ transformed intensities of Affymetrix microarray data and relative expression measures by qRT-PCR.

Gene of Interest	Microarray versus qRT-PCR	P-value
Sb01g017270	0.767	0.001
Calculated using cor.test in R		

Figure B.3 Pearson's correlation between expression levels determined by microarray and qRT-PCR. * The mRNA levels for Sb01g17270 were determined by qRT-PCR and correlated with RMA normalized microarray expression scores. All data were $\log_2$

Figure B.4 Number of genes expressed in each of the 78 samples. Total: number of gene expressed in at least one organ (19,354; 70% of all genes on the array). Common: genes expressed in all 78 tissue types (4526; 15% of all genes on the array)

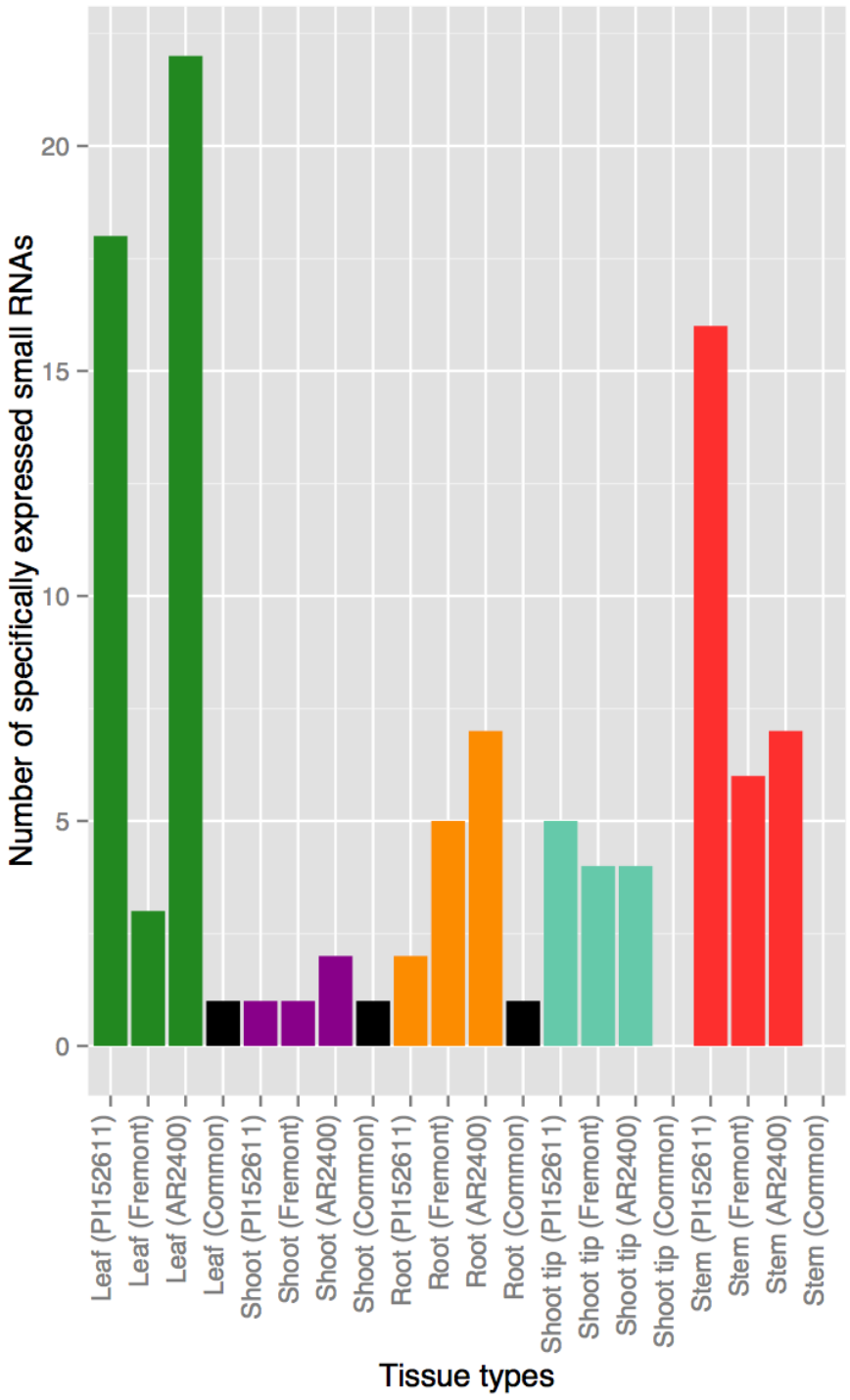


Figure B.5 Number of tissue-specific small RNAs across sorghum ideotypes. AR2400: biomass sorghum; Fremont: sweet sorghum; PI152611: forage sorghum; Common: number of genes in common among all three ideotypes

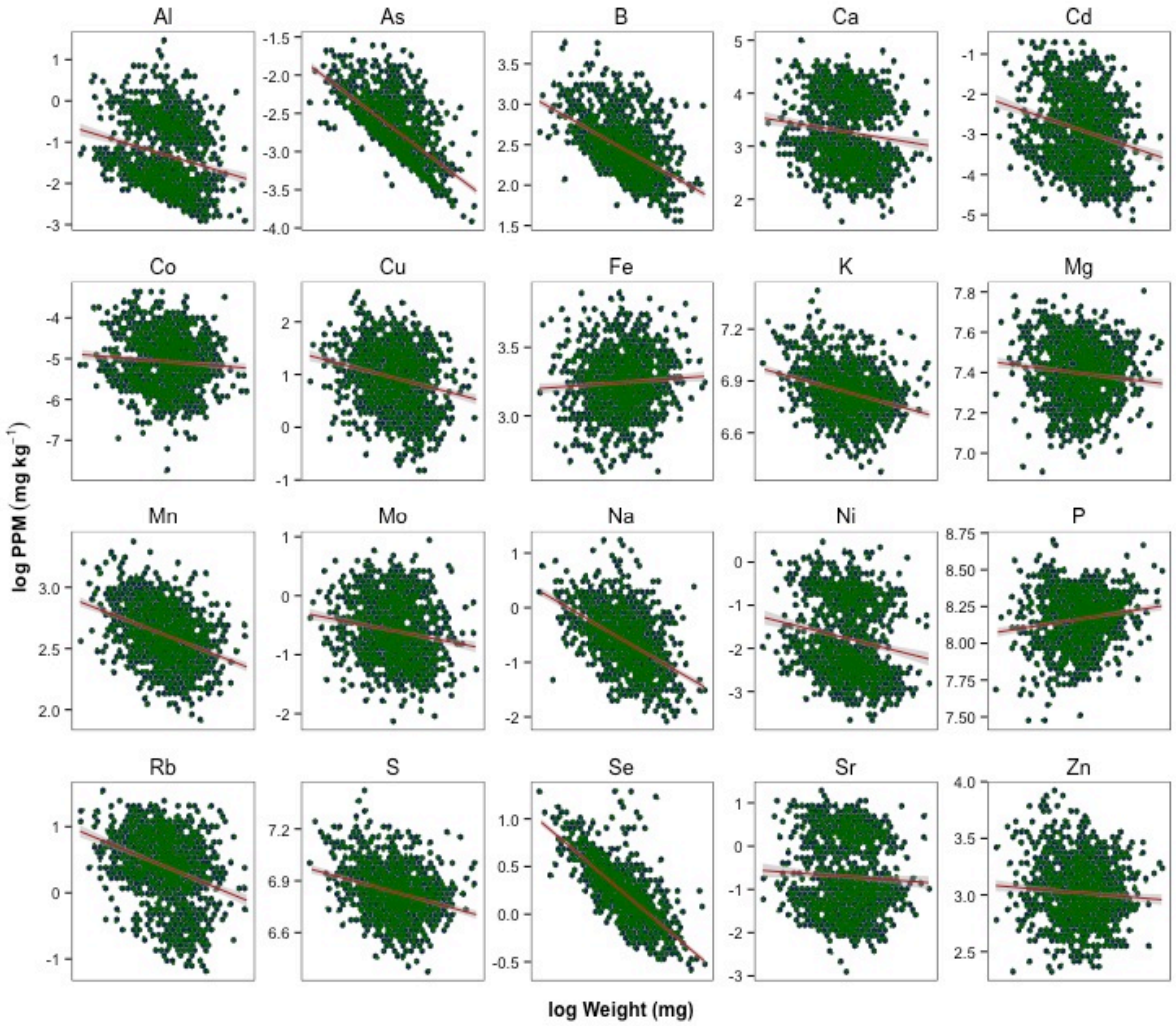


Figure B.6 Scatterplots depicting the correlation between log normalized seed weight and elemental concentration in all accessions from SAP and CHP association panels (SAP 2008, SAP 2012, SAP 2013 and CHP 2010)

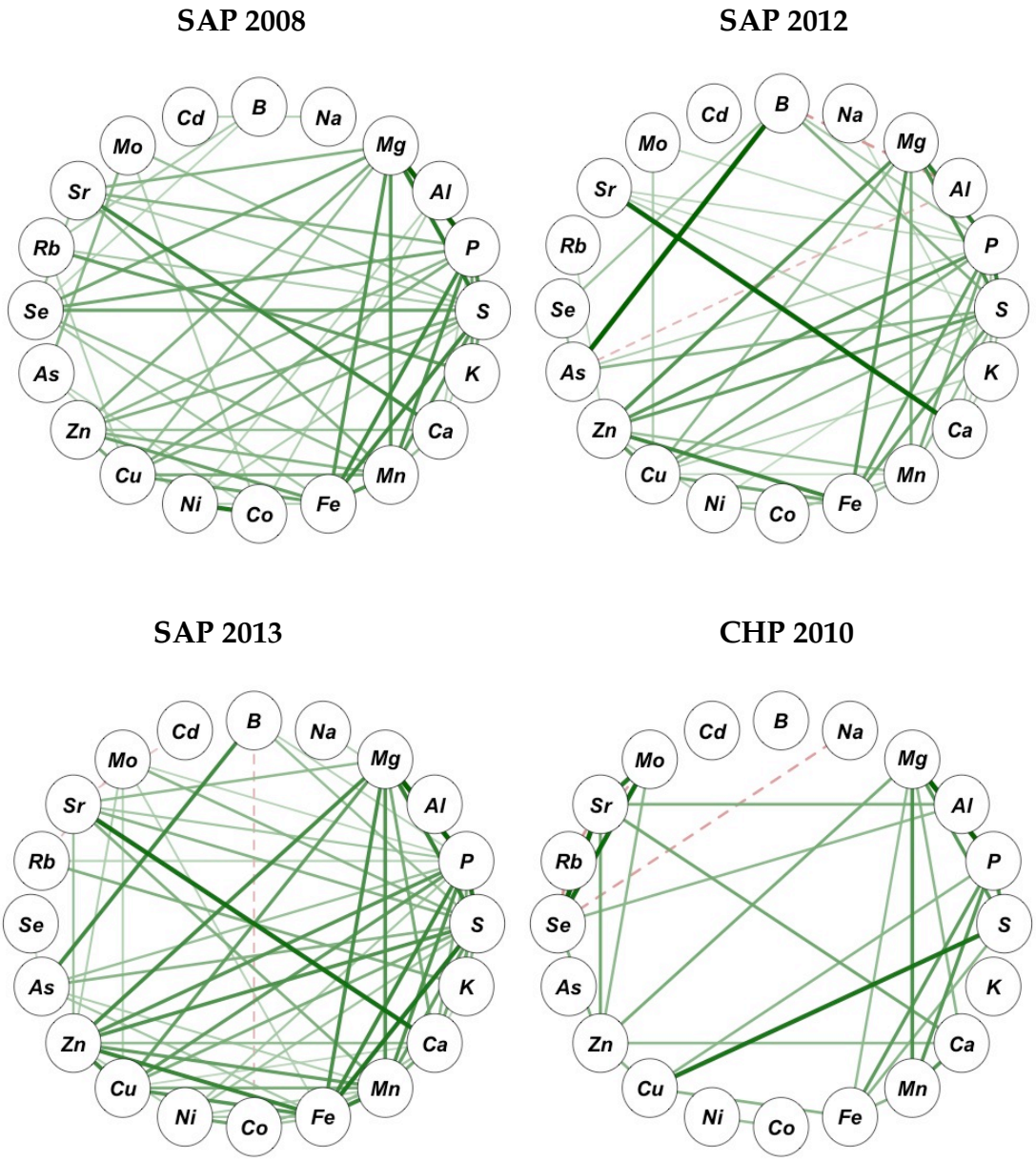
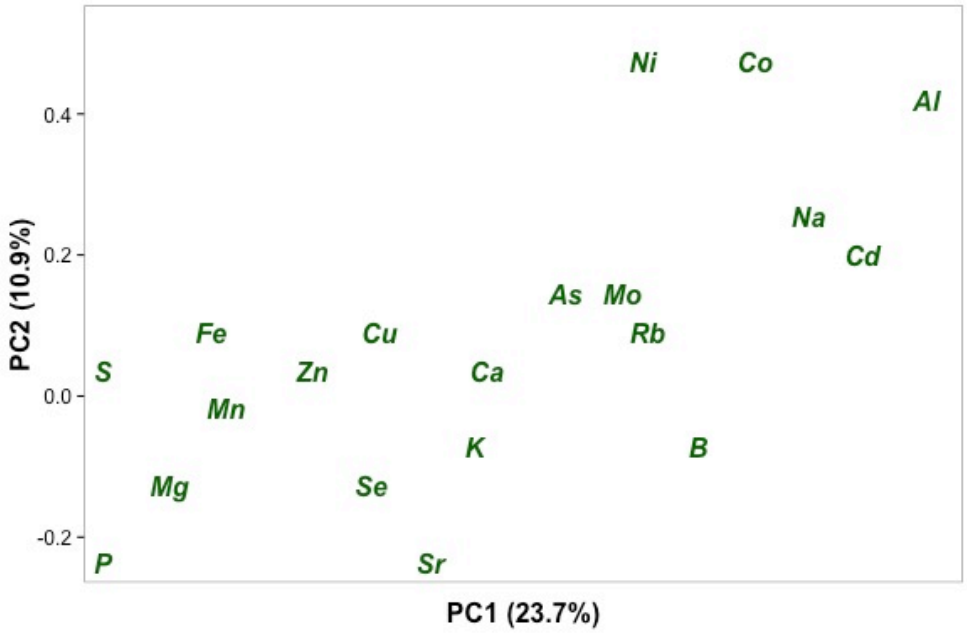


Figure B.7 Correlation network of seed mineral concentrations calculated across replicates from SAP and CHP association panels. Green solid lines represent positive correlation values. Red dashed lines represent negative correlation values. Intensity and thickness of lines indicate degree of correlation. Mineral correlation values can be found in Table A.12

SAP 2008



SAP 2012

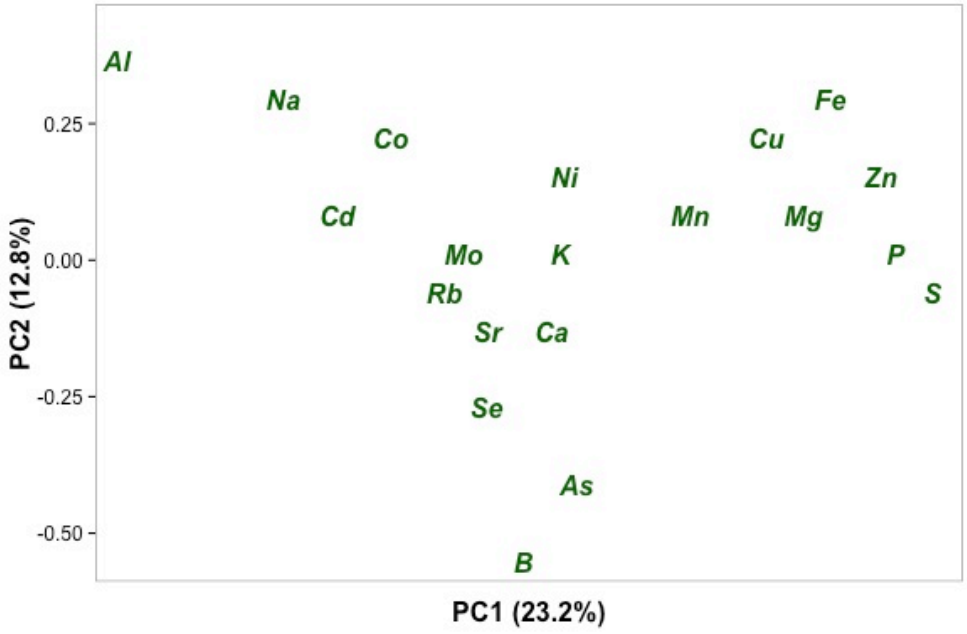
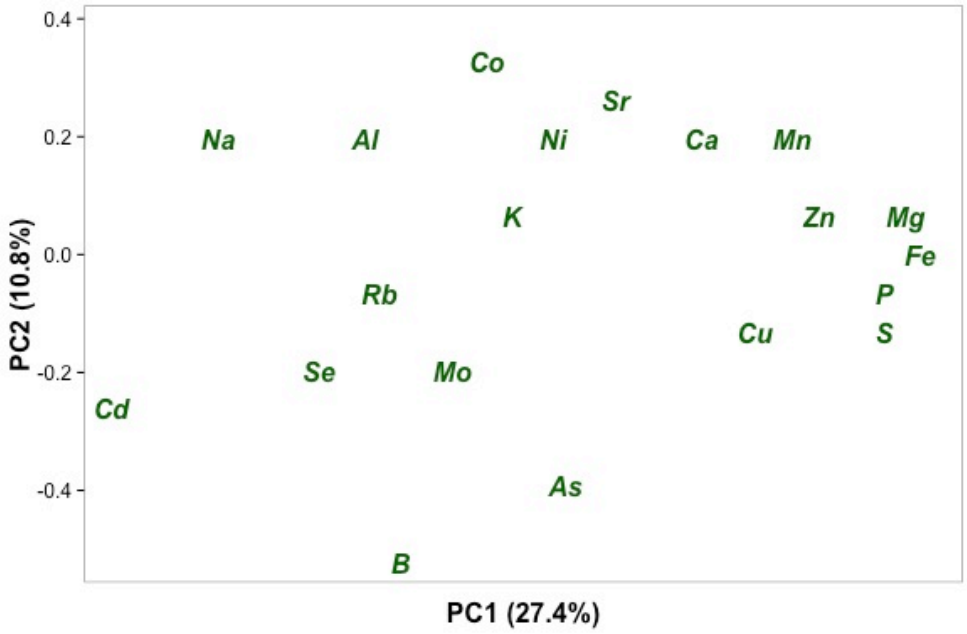


Figure B.8 Principal component analysis applied to the average seed concentrations for 20 minerals in the SAP and CHP association panels. Each symbol represents a single element

SAP 2013



CHP 2010

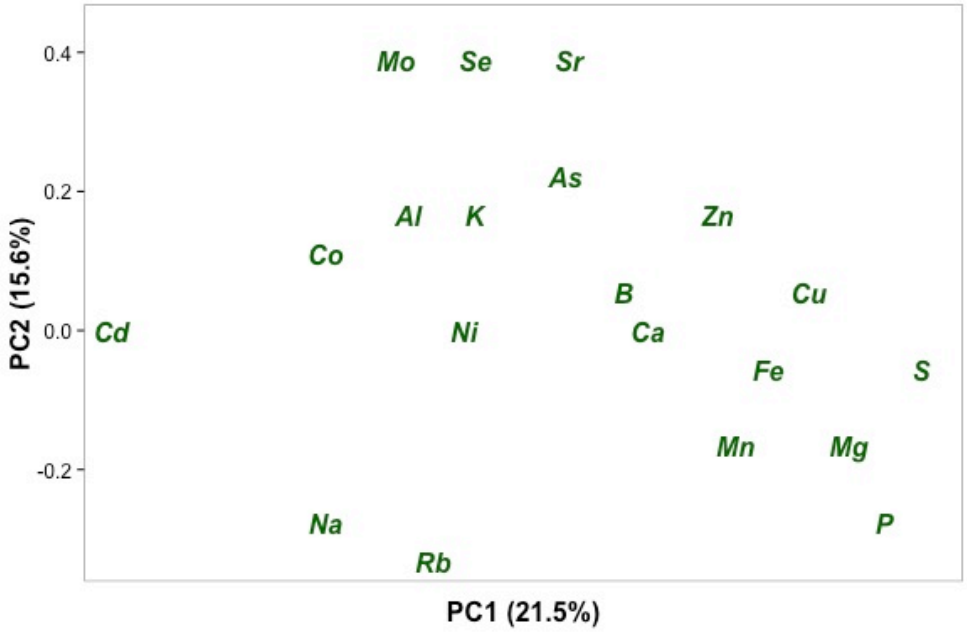


Figure B.8 (cont.) Principal component analysis applied to the average seed concentrations for 20 minerals in the SAP and CHP association panels. Each symbol represents a single element

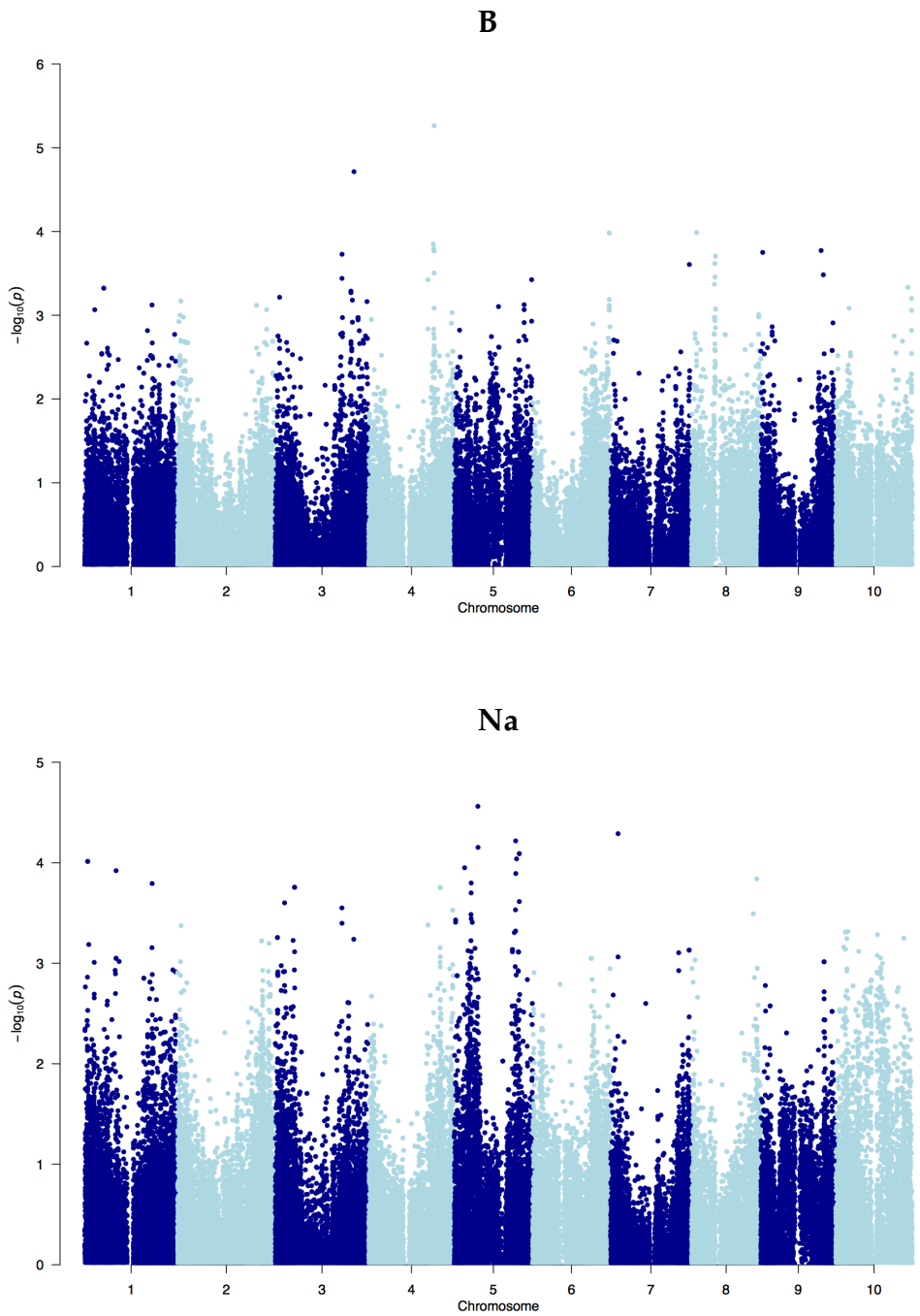


Figure B.9 Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

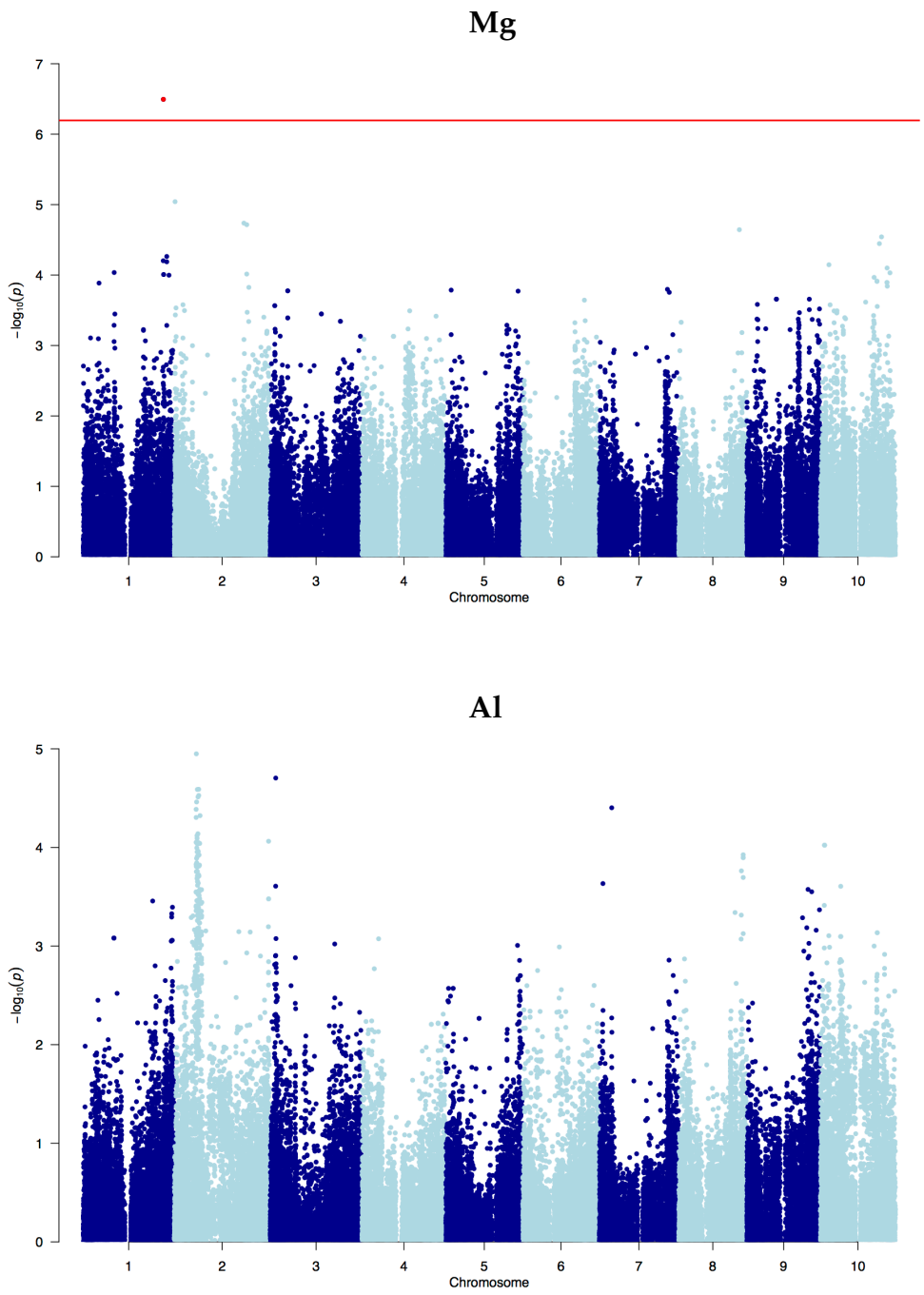


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

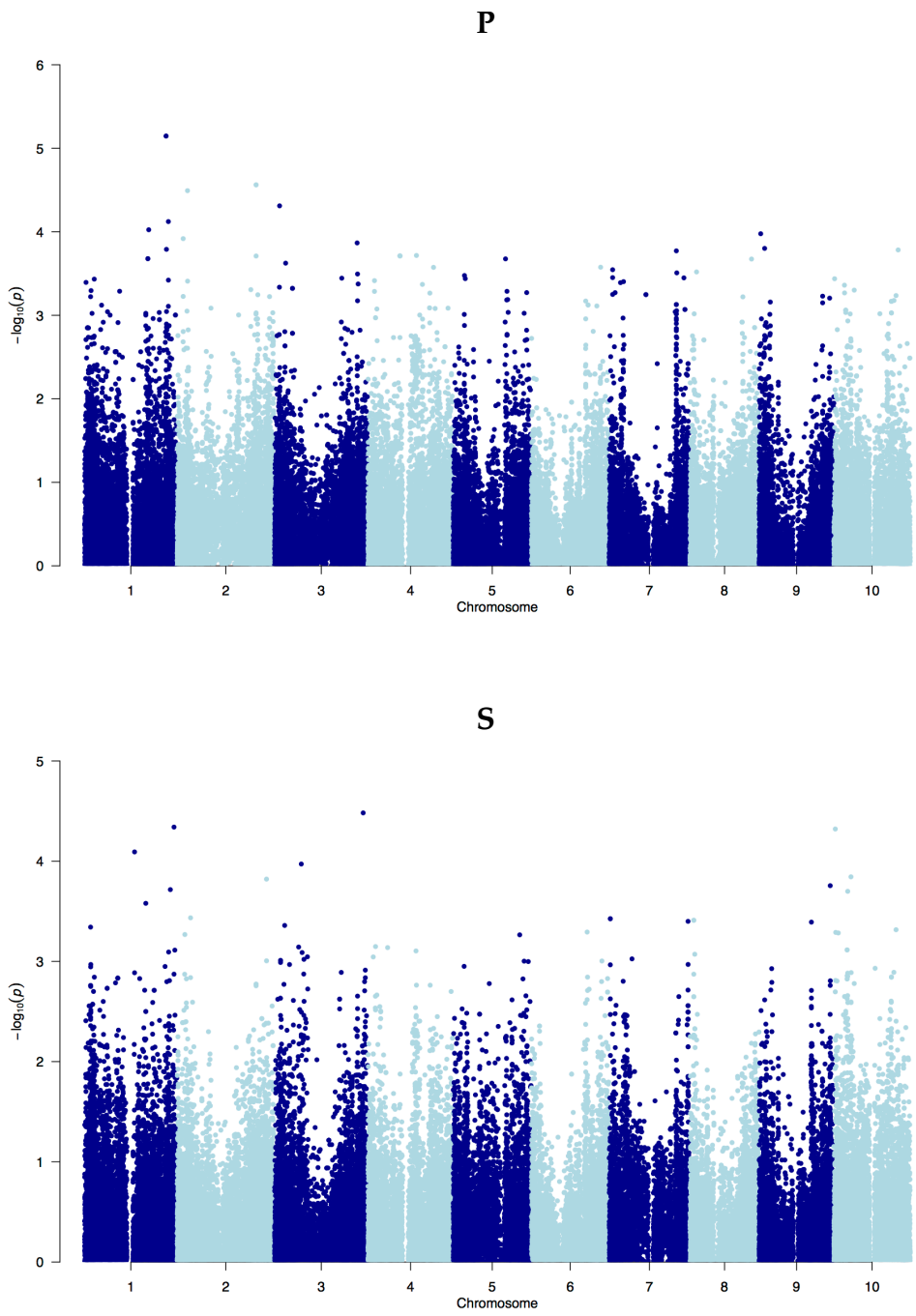


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

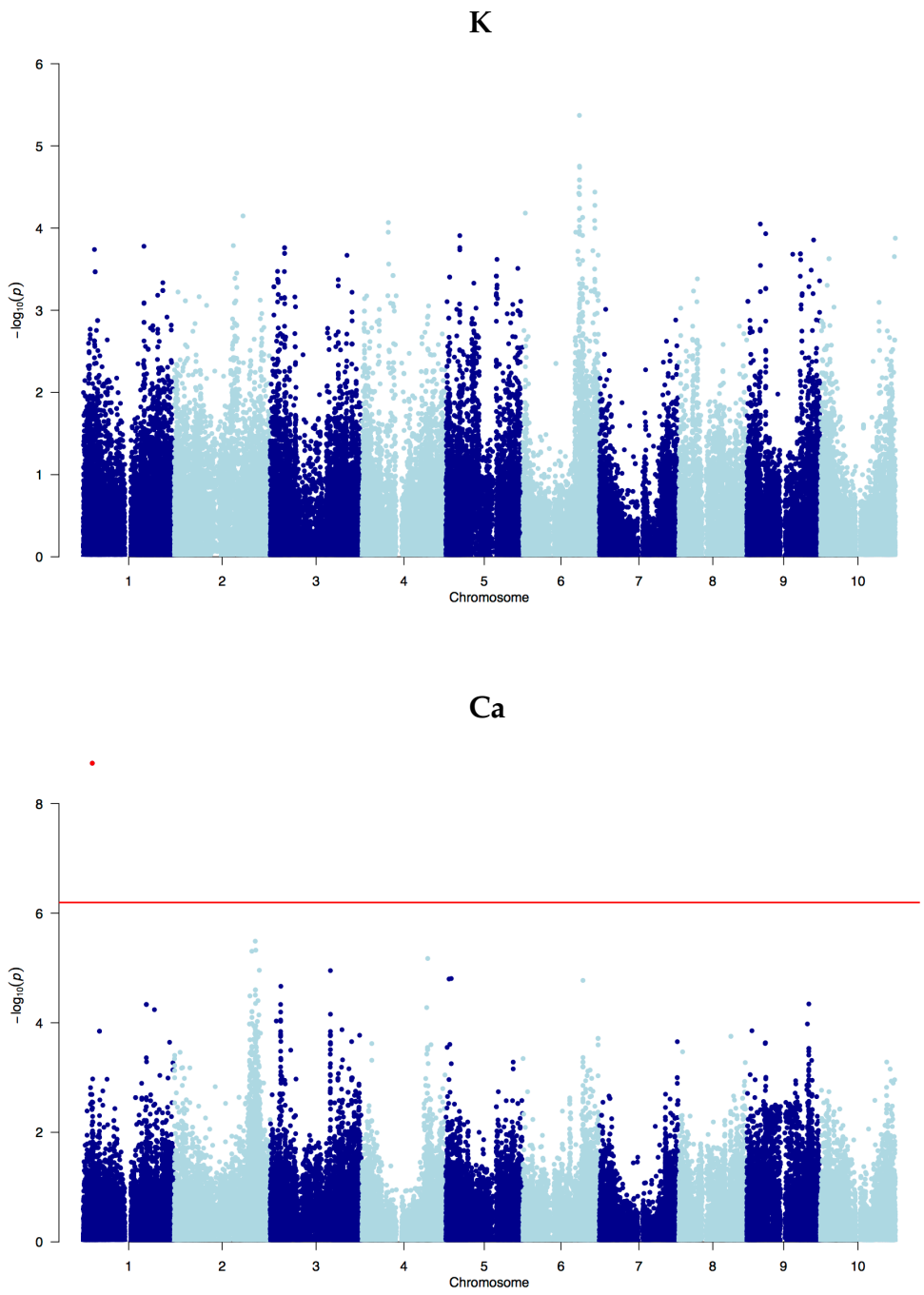


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

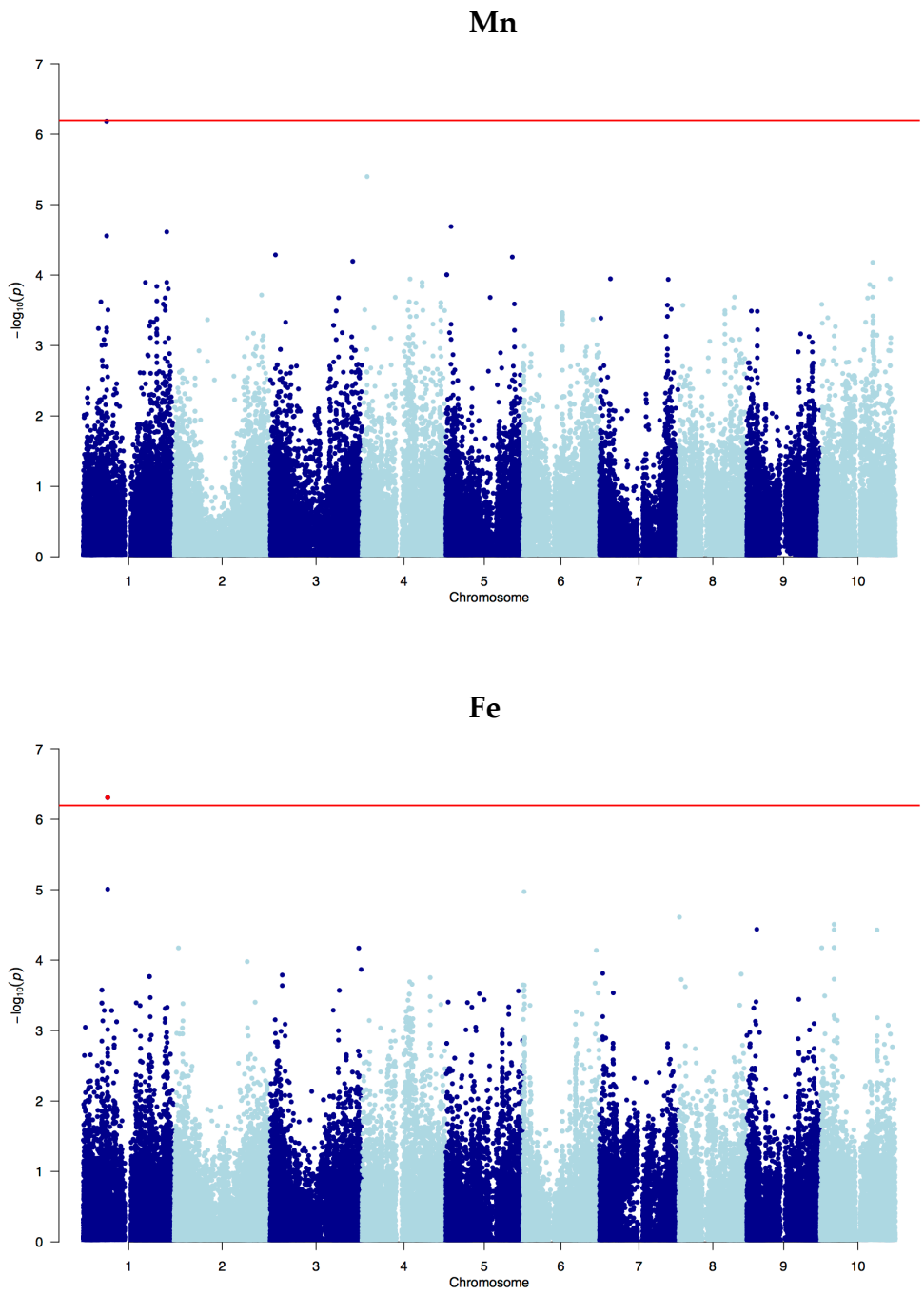


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

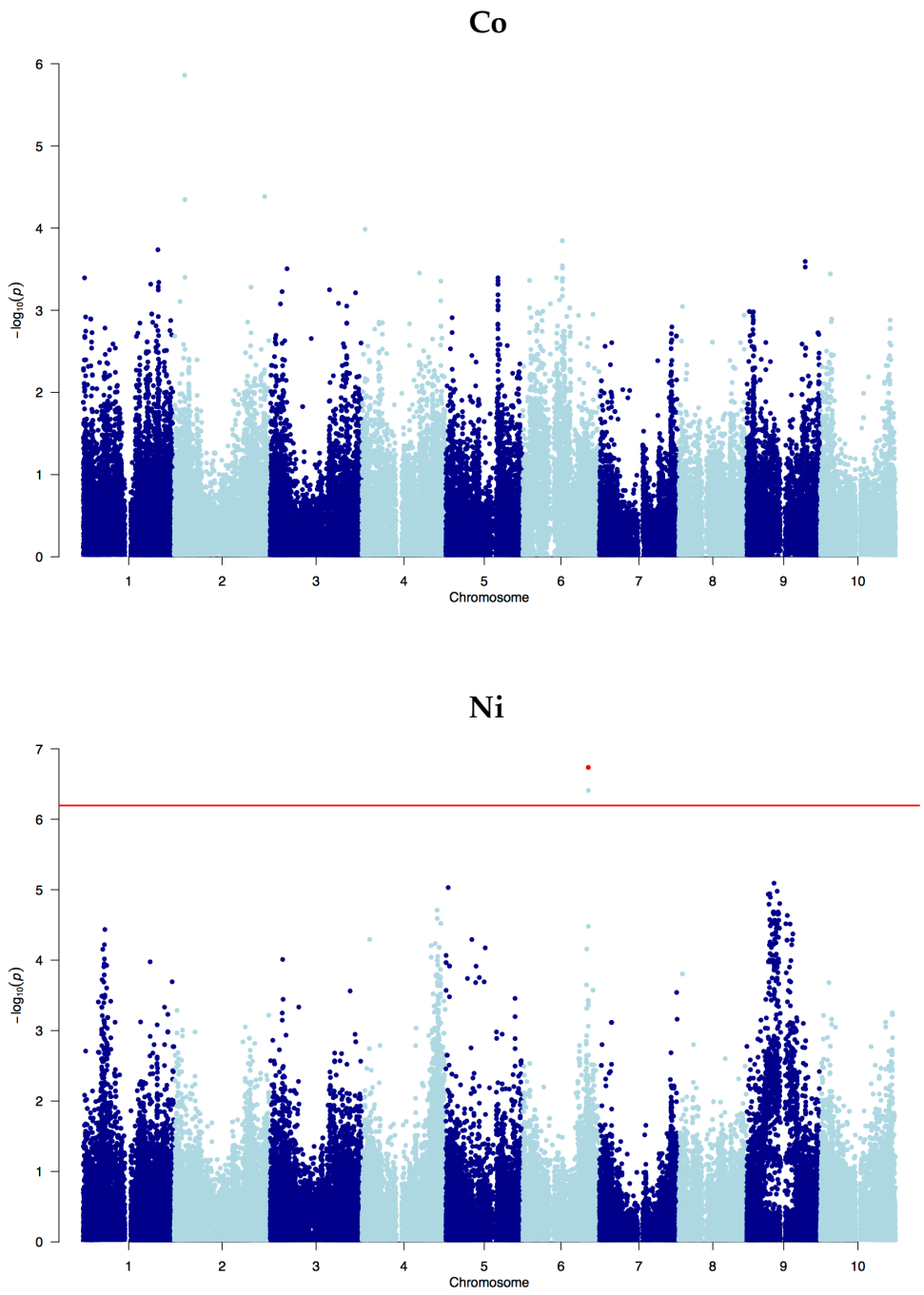


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

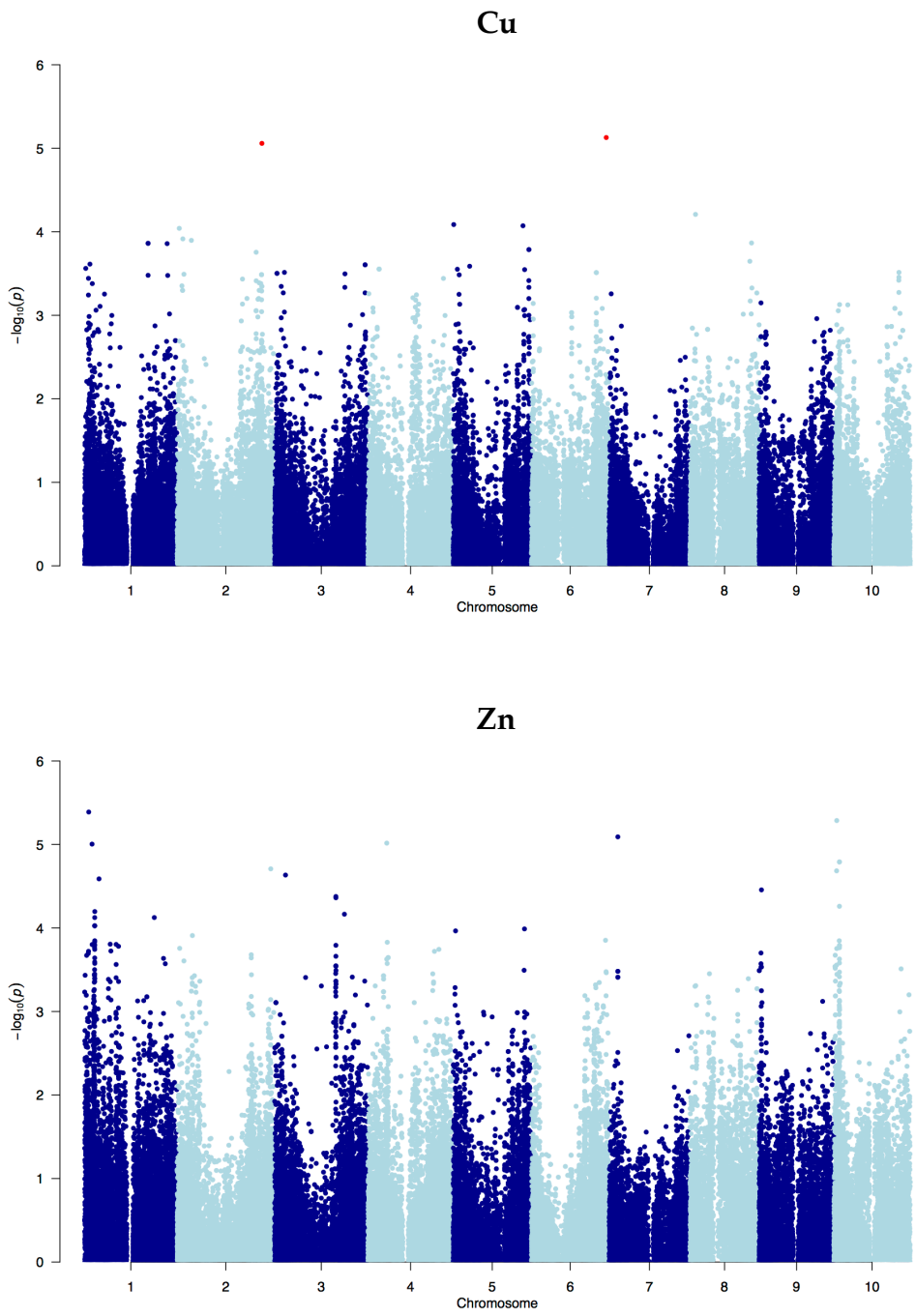


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

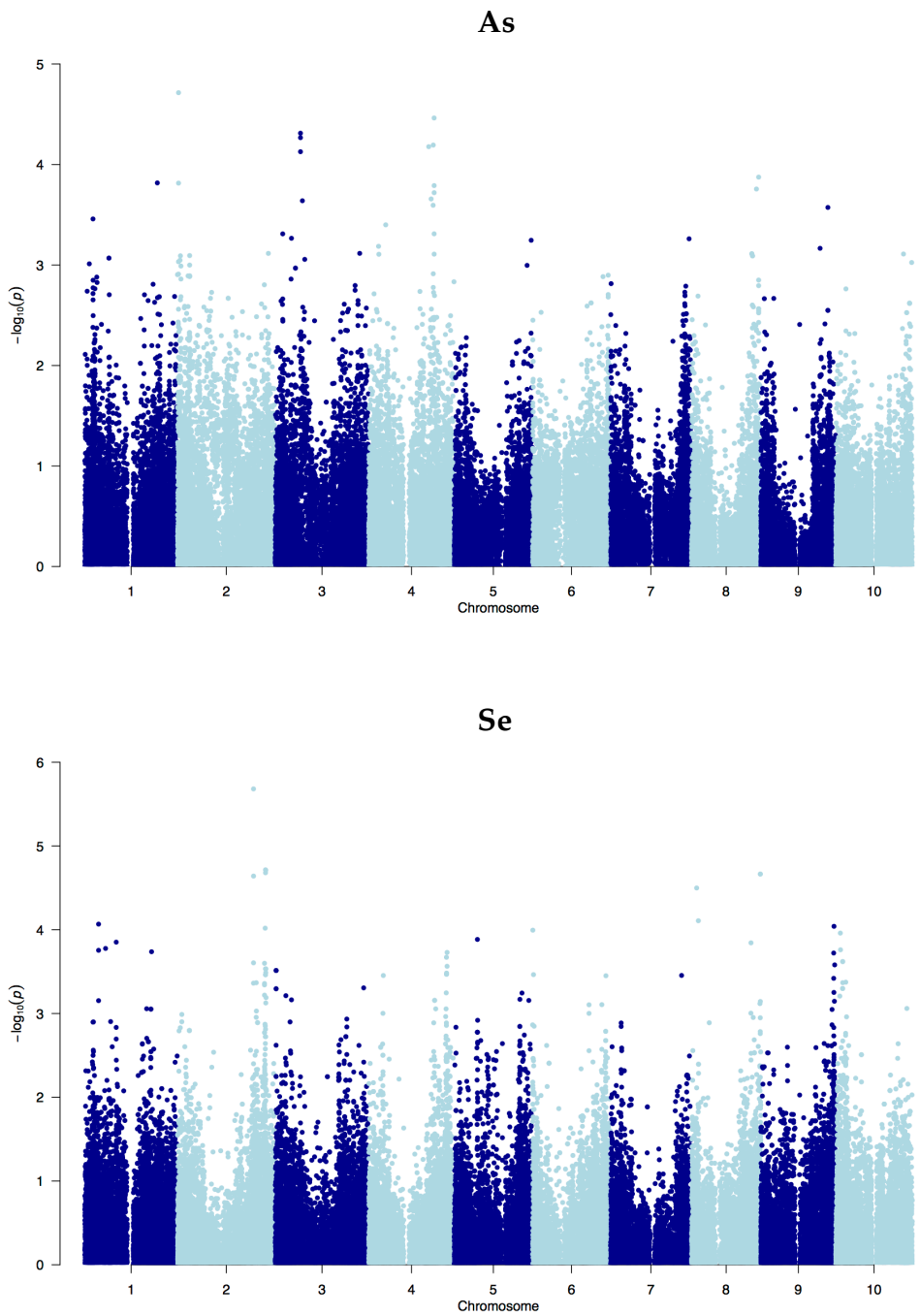


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

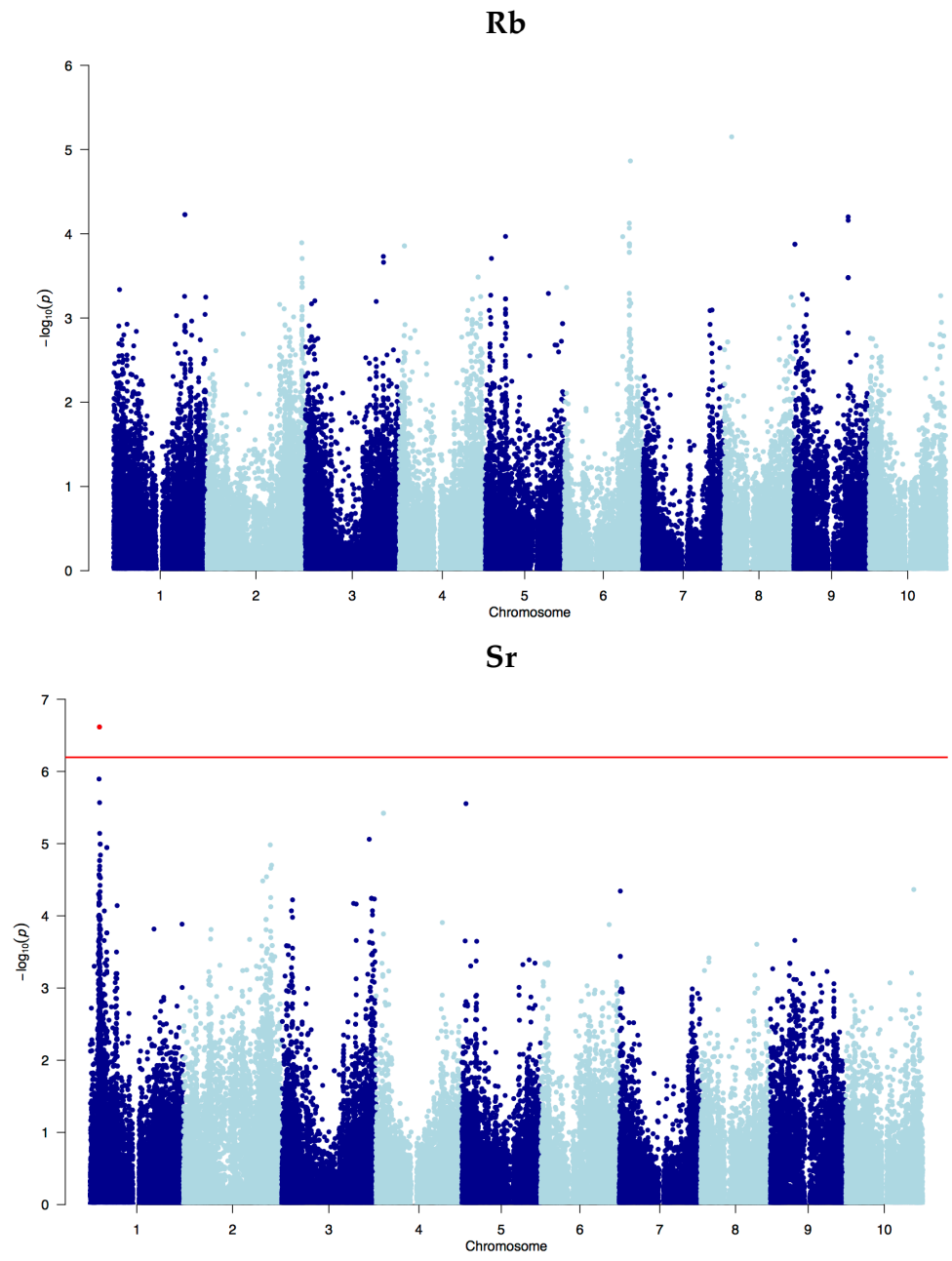


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

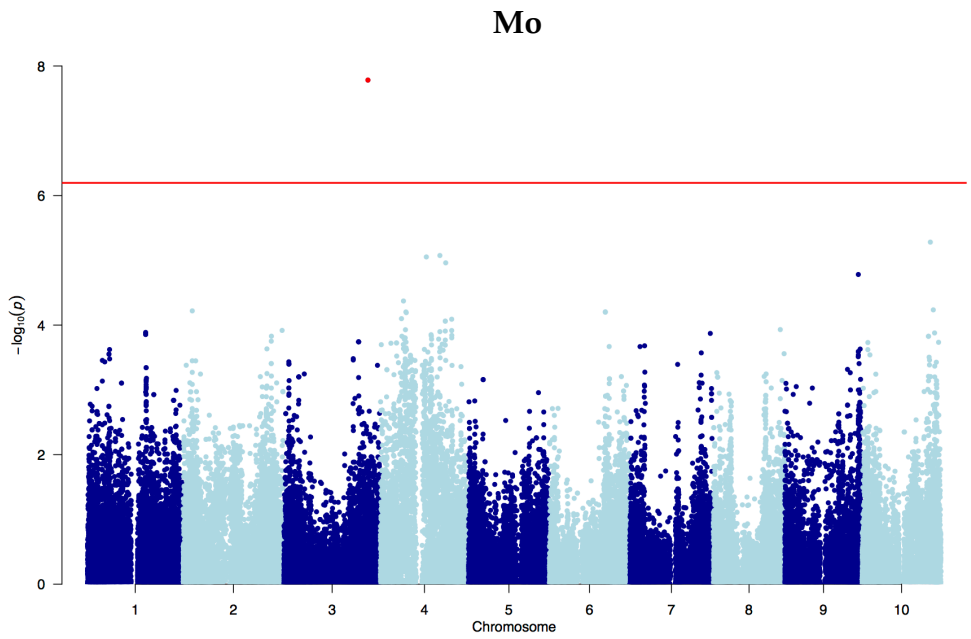


Figure B.9 (cont.) Manhattan plots for 19 minerals displaying GWAS results ($-\log_{10}(P)$) for the 10 sorghum chromosomes (x-axis) and associated P values for each marker (y-axis)

APPENDIX C

COPYRIGHT PERMISSION

Free articles

All articles in BioMed Central journals are available online without charge or other barriers to access. The following journals have published a small number of articles that, while freely accessible, are not open access as outlined in the section above:

Alzheimer's Research & Therapy
Arthritis Research & Therapy
Breast Cancer Research
Critical Care
Genome Biology
Genome Medicine
Stem Cell Research & Therapy

These articles may be flagged as 'Free' content and their bibliographic information may indicate that copyright rests with the publisher or another organization. With the exception of a few co-published articles (see below), these free articles may be reproduced for non-commercial purposes without formal permission from BioMed Central or payment of fees, provided full attribution is given. If you wish to reproduce such an article, or parts thereof, for commercial purposes (other than original figures and/or tables), please contact us to check whether formal permission is needed.

Co-publications: Different rules may apply to articles which are co-published (i.e. published in a BioMed Central journal and simultaneously in another publication); in such cases a statement of co-publication is included in the bibliographic information. Please contact us for permission to reproduce content from these articles.