Spring 5-15-2019

A Combinatorial Approach of Ionomics, Quantitative Trait Locus Mapping, and Transcriptome Analysis to Characterize Element Homeostasis in Maize

Alexandra Asaro
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Computational and Systems Biology

Dissertation Examination Committee:
Ivan Baxter Chair
Barak Cohen, Co-Chair
Heather Lawson
Blake Meyers
Kenneth Olsen
Chris Topp

A Combinatorial Approach of Ionomics, Quantitative Trait Locus Mapping, and Transcriptome Analysis to Characterize Element Homeostasis in Maize

by

Alexandra Asaro Fikas

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2019
St. Louis, Missouri

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA - Analysis of variance

eQTL - Expression quantitative trait locus/loci

DNA - Deoxyribonucleic acid

FDR - False discovery rate

GDD - Growing degree days

GO - Gene ontology

GWAS - Genome-wide association study

G x E - Genotype by Environment

IBM - Intermated B73 x Mo17

ICP-MS - Inductively-coupled plasma mass spectrometry

LDA - Linear discriminant analysis

LOD - Logarithm of the odds

MCL - Markov cluster

PCA - Principal components analysis

PC - Principal component

QTL - Quantitative trait locus/loci

QEI - QTL by environment interaction

RIL - Recombinant inbred line

RNA - Ribonucleic acid

rRNA - Ribosomal RNA

RNA-seq - RNA sequencing

RPKM - Reads per kilobase of transcript, per million mapped reads

SNP - Single nucleotide polymorphism

ACKNOWLEDGEMENTS

Several people have made immeasurable contributions to my graduate school experience. First and foremost, my advisor, Ivan Baxter, who has been an encouraging and supportive mentor while also giving me the freedom to grow as a scientist. If not for Ivan, I doubt I would have adventured into the world of computational biology, and I am enormously thankful that he believed in me from the beginning. I am also grateful for the support of my fellow lab members. I sincerely appreciate all of the valuable perspectives offered to me by members of my lab, the Topp lab, and others at the Danforth Center, as well as by our collaborators, especially Brian Dilkes, who contributed an extraordinary amount of insight on complex problems. I thank the members of my thesis committee for their time and assistance during this process. Much of this work was funded by the William H. Danforth Plant Science Fellowship and the USDA National Institute of Food and Agriculture Predoctoral Fellowship.

On a personal note, I'd like to acknowledge my parents and sister for their unending support. My sister has been the friend I needed throughout graduate school and my parents gave me confidence that I could achieve anything I set my mind to. Finally, I want to offer special thanks to my husband, Giannis, who came into my life at one of the lower points along my graduate school trajectory. He motivated and inspired me to work hard and persevere for this degree and I would not be where I am today without him.

Ally Asaro Fikas

Washington University in St. Louis
May 2019

Dedicated to my dog, Bentley, who has made me smile every day along this journey.

ABSTRACT OF THE DISSERTATION

A Combinatorial Approach of Ionomics, Quantitative Trait Locus Mapping, and Transcriptome Analysis to Characterize Element Homeostasis in Maize

by

Alexandra Asaro Fikas

Doctor of Philosophy in Biology and Biomedical Sciences

Computational and Systems Biology

Washington University in St. Louis, 2019

Dr. Barak Cohen, Chair

Dr. Ivan Baxter, Co-Chair

In plant systems, genetic and biochemical pathways impact uptake of elements from the soil. These environment-sensitive pathways often act in the root tissue to impact element concentrations throughout the plant. In order to characterize element regulation as well as apply ionomics to understand plant adaptation, perspectives are needed from multiple tissues and environments and from approaches that take interactions between elements into account. The work described in this thesis includes multi-environment and multi-tissue experiments that connect variation in genetic sequence, and in gene expression, with variation in element accumulation. The associations found here include those that are sensitive to environment, reflecting the complex environmental influence on the ionome, as well as those that exhibit consistent effects across different environments. A variety of statistical tools were employed to model genetic by environment interactions and test methodologies that can be applied to future studies of the ionome with more in-depth environmental data. Genetic loci with strong effects on

elements across environments were further explored using root-based gene expression data, which identified candidate genes and gene networks underlying element accumulation. Additional research on these candidate genes has the potential to improve our understanding of the genetic basis of homeostatic processes that involve the ionome, as well as isolate targets for genetic modification or selective breeding that can enhance nutritional content and adaptive capacity of crops.

CHAPTER 1:
INTRODUCTION

OVERVIEW OF IONOMICS

Mineral nutrients play key roles in cellular processes as cofactors in biochemical reactions, structural components, and electrochemical regulators. Plants maintain ion homeostasis via complex regulatory systems sensitive to both environmental and physiological changes. The term *ionomics* was coined by Lahner et al. in 2003, in the first high-throughput elemental profiling study, performed in the model plant species Arabidopsis [1]. Elemental profiles, typically measured in seeds or leaves using high-throughput inductively-coupled plasma mass spectrometry (ICP-MS), reflect the genetic by environment interactions that influence nutrient content throughout the plant. Element content is a complex genetic trait; it does not follow standard Mendelian patterns of inheritance, but instead is quantitative in nature, determined by multiple genes and gene interactions. Elemental profiling of genetically distinct plants can be used in quantitative analyses to isolate regions of the genome controlling element accumulation. Because seeds and leaves represent a lifetime of nutrient accumulation, genetic variation that impacts processes throughout the plant, such as those occurring in the root, will be reflected in seed or leaf profiles.

Elemental signatures provide a means to characterize difficult-to-measure phenotypes and diagnose stress responses. The sensitivity of the ionome to environmental and physiological states encoded by the genome renders ionomics a useful tool for understanding not only element

regulation but also related traits involved in plant adaptation to the environment, a major focus of basic plant science research and current global agriculture improvement efforts. For instance, an early study of ionome mutants revealed a mutant with both an altered shoot elemental profile and root-dependent increased water stress tolerance [2]. By using the relationship between shoot element concentrations and root solute transport, this elemental profiling screen identified a genetic variant with increased vigor in drought conditions. The advantages of using ionomics as a proxy for plant adaptation include the relatively high heritability of seed and leaf mineral nutrient content, particularly for certain elements, and the ability to discern element concentrations with comparatively low cost in a high-throughput pipeline.

Since the introduction of ionomics, researchers have used high-throughput elemental profiling and quantitative genetics to uncover loci and genes underlying element homeostasis. However, the majority of loci have not been resolved to genes and the functions of genes that have been discovered and their roles within genetic networks remain largely unknown. Sources beyond traditional genetics are required to understand elemental profiles, which, like most complex traits, exhibit high genetic by environment interaction. Furthermore, elements do not behave independently, but are interrelated, with factors affecting multiple elements. Genetic regulation of the ionome can be best characterized by going beyond single-element and single-tissue approaches, employing multivariate analysis and relating root gene expression to the whole plant ionome. This combination of techniques will improve resolution of quantitative trait loci (QTL) to genes and identify gene networks influencing the ionome. Such information can be applied in breeding programs focused on improving crop nutrition or eliminating accumulation of toxic micronutrients, such as cadmium, from food sources. By identifying genes and contextualizing them in networks, we can link genes that regulate the ionome to adaptive

processes, such as drought tolerance, and thereby provide a means to profile for genotypes that thrive in extreme environments.

HISTORY OF IONOMICS

The first ionomics study was conducted by Lahner et al. in Arabidopsis [1]. The group isolated mutants with altered leaf element profiles by using ICP-MS for elemental profiling of 18 elements and a forward genetics approach. A key result of this study was that the majority of ionome mutants exhibited altered profiles of multiple elements; only 11% of ionome mutants had significant changes in a single element. The high incidence of multi-element changes provided initial evidence that the ionome functions as a network. Furthermore, multi-element profiles were capable of distinguishing groups of mutants through linear discriminant analysis. Several subsequent experiments furthered the ionome as a network hypothesis; predictive multi-element signatures for iron (Fe) and phosphorous (P) statuses were identified in Arabidopsis [3], multi-element variation was described using principal components analysis (PCA) in the model legume *Lotus japonicus* [4], and correlations between calcium (Ca) and magnesium (Mg) were observed in *Brassica oleracea* [5]. These findings suggest that genes regulating multiple elements comprise a large fraction of element homeostasis networks, highlighting the need to develop methods capable of detecting such components [6].

Forward genetics allowed for characterization of informative mutants in Arabidopsis and other organisms, such as *Lotus japonicus* [7] and soybean [8]. The first ionomic mutant cloned in Arabidopsis was the *enhanced suberin1-1* (*esb1-1*) mutant [2], a mutant with a multi-element leaf ionome signature. The elemental changes observed in the *esb1-1* mutant were attributed to aberrant lignin and suberin deposition in the Casparian strip, a structure that acts in roots as a

selective filter for water and nutrient transport [9]. Additional Casparian strip-related mutants have been identified in Arabidopsis, all showing a multi-element phenotype [9, 10].

Natural variation has also been a robust resource of genetic variants involved in element homeostasis. In Arabidopsis, natural accessions were used to identify a novel allele in the Na transporter *AtHKT1* that reduces *AtHKT1* expression in the roots and subsequently increases Na in shoots [11]. This study was one of the first to show that root gene expression can be a strong determinant of element content throughout the whole plant. Baxter et al. also used natural variation in Arabidopsis to identify the molybdenum transporter gene *MOT1* and the causal deletion in the gene. Similar to the mechanism of the *AtHKT1* Na transporter variant, the deletion in *MOT1* was shown to cause depletion of whole-plant Mo through reduction of gene expression in the root [12]. An additional MOT1 variant was found to be correlated with Mo content of soils and allelic variation in MOT1 was associated with adaptation to native soil type [13]. Similar studies in Arabidopsis identified ferroportin mutants with aberrant Fe and cobalt (Co) localization [14] and characterized of leaf sulfate QTL at the genes *APR2* and *ATPS1*, which encode enzymes belonging to the same sulfate accumulation pathway [15, 16].

QTL mapping and genome wide association studies (GWAS) are powerful statistical methods that can connect natural variation with specific phenotypes of interest. These methods have been used to find and describe loci impacting kernel and leaf ionome traits. For example, in Arabidopsis, GWAS on leaf ionome variation combined with transgenic complementation allowed Chen et al. to describe a polymorphism in the heavy metal ATPase gene *HMA3* that decreases leaf Cd [17] and identify a new arsenate reductase enzyme, HAC1, with a key role in As reduction [18]. These methods have been applied in a variety of species, including maize [19, 20], rice [21], sorghum [22], and soybean [23]. QTL mapping is typically carried out in a bi-

parental population whereas GWAS can include multiple parents. While QTL studies often exhibit increased power and can identify loci of interest, bi-parental populations offer a limited amount of genetic variation to study and genetic resolution can be relatively low. GWA studies benefit from expanded genetic variation and increased genetic resolution which allows for easier association of loci with specific genes [24].

Many experiments using genetic analyses have also implemented techniques like expression quantification and imaging to further describe mutant phenotypes and genes known to be involved in element homeostasis. Expression of potassium channel genes in Arabidopsis was tracked using real-time PCR [25]. Fluorescence imaging was used to localize the MOT1 molybdenum transporter to the mitochondria [12] and the NaKR1 metal-binding protein to the companion cells of the phloem [26]. Grafting has been used repeatedly to determine root-based sources of variation in shoot and leaf ionome mutants [12, 25, 26]. DNA sequencing of pooled mutants followed by microarray analysis isolated the causal gene in *myb36-1*, a mutant exhibiting a multi-element phenotype similar to that of other Casparian strip mutants [9]. Expression analysis and visualization techniques were then applied to determine the impact of mutant MYB36 on target gene expression and characterize cell type localization of the mutant protein and its associated targets.

The ionome is highly responsive to the environment and the genetic mechanisms influencing the ionome can vary depending on environment. Previous investigations have looked at the relationship between the environment and the ionome. These include surveys of the Arabidopsis leaf ionome under varying soil salinity [27] and the tomato leaf ionome during water stress conditions [28]. While work has been done to characterize QTL by environment interactions underlying ionomic variation [20, 29–32], orthogonal datasets and G x E models that

include specific environmental variables are needed to obtain a gene-level understanding of these interactions.

IMPORTANCE AND APPLICATONS OF IONOMICS

The study of plant adaptation using ionomics can be applied to address numerous pertinent issues relating to abiotic environmental factors, crop yield, and sustainable agriculture. Most essential elements in plants, excluding carbon and oxygen, are derived from the soil. Because plants need to adapt their mechanisms for acquisition, transport, and storage of mineral nutrients to specific soil conditions and environmental changes, the underlying systems are flexible, with substantial variation among genotypes [33]. Regulation often involves multiple elements at once via processes such as co-transportation, as seen with Fe and Zn [34, 35], Ca and Mg [5], and Na and K [26]. Chemical analogs, such as Ca and Sr or K and rubidium (Rb), frequently display similar ionomic profiles [6].

Abiotic factors, such as non-ideal soil nutrient levels and harsh environmental conditions, pose a threat to crops unable to adapt to such stressors. High levels of certain elements in soil can be toxic to the plant and/or consumers. Soil element concentration, drought, salinity, and invasive species all effect plant growth dramatically and vary across environments. In order to develop crops able to flourish in particular conditions or maintain crop improvements across diverse environments, we must understand the genetic by environment interactions that underlie specific adaptive mechanisms. The growing human population demands yield and nutrition improvement in crops, with nutrient deficiencies being a widespread current issue, especially in areas of poverty. Yield needs to grow at an exponential rate parallel to that of population growth in order to provide required food and biofuels. Due to global climate change, this yield

acceleration must occur in growth environments that are predicted to become increasingly hostile or unfit for current agricultural practices [36]. Nutrient deficiencies pose serious health risks worldwide, particularly in developing nations [37, 38], making it necessary to increase not only the quantity of crops produced but also their nutrient value. Sustainable methods are needed to minimize waste and pollution, maximize water-use efficiency and soil nutrients, and prevent destruction or contamination of local ecosystems. These goals can be promoted by breeding or engineering crops that achieve optimal mineral nutrient homeostasis in extreme conditions without requiring environmentally damaging interventions, such as extensive fertilizer application or irrigation.

Several previous ionomics studies have informed on ion homeostasis and plant adaptation with results applicable to addressing abiotic stressors, improving yield, and/or increasing sustainability. For example, analyses have mapped QTL associated with low-phosphate tolerance, many of which are also involved in root traits such as root length and root hair density [39]. The use of ionomics and other –omics approaches identified transporters and other factors involved in P homeostasis that were used for transgenic manipulation. Overexpression of Pi-regulating factors that respond to Pi deficiency altered traits desirable for uptake, including root morphology, increased expression of Pi transporters, and conferred low Pi tolerance without inducing Pi toxicity [40–42]. Network construction that accounts for expression variation across different tissues will aid in choosing genes that can be modified to improve abiotic conditions such as low Pi availability without fitness-decreasing consequences. Important micronutrients for human nutrition have been studied with ionomics. The IRT1 iron transporter was found to play a role in the iron deficiency response through gene expression changes and concurrent increases in other metals. by transporting additional metals [43]. Other studies in have identified loci and

QTL by environment interactions underlying leaf and grain concentration of toxic elements, such as Cd [17, 20]. QTL regulating the ionome under drought stress have been mapped in Arabidopsis [44] and rice [45]. Pathogen response has also been explored in Arabidopsis, with a gene related to pathogen response linked to leaf potassium homeostasis [25]. The potential for temperature changes to alter the ionome was evaluated in Lotus japonicus, with significant alterations in the shoot ionome observed in reaction to sub-optimal root zone temperatures [46].

Analysis of micro- and macro- mineral nutrients in the seeds or leaves of a plant along with genotypes, environmental variables, and other phenotypes such as height, biomass, gene expression profiles, or metabolite panels, has the potential to relate genes and gene networks that control the ionome to developmental state and environment. By identifying favorable allele-environment pairs we can tailor agricultural practices to our specific needs. Unlike the practice of random breeding for beneficial alleles over several generations, the application of information gained from high-throughput genetic and phenotypic studies can produce efficient, targeted changes in plant adaptive capacity.

THE FUTURE OF IONOMICS

Past studies in the area of ionomics provide ample evidence that ionomics is a valuable tool for understanding the genetic basis of ion homeostasis and plant environmental adaptation. While some genetic variants have been characterized in-depth, with insight into the causal gene mutation and functional basis of the mutant phenotype, the specific functions of the majority of mapped loci remain unclear. Genes that have been identified have often not yet been placed in the context of genetic networks. The application of ionomics to agriculture while avoiding unforeseen side-effects of genetic modifications or unfavorable allele-environment combinations

demands an extensive understanding of gene-gene interactions and the interaction of environment with genes and genetic networks.

Studies on genetic by environment interactions controlling the ionome will be improved by data collection across more environments with more extensive metadata. Comparative analyses are often complicated by studies often having limited data on field growth environment and varying growing practices. Uniform practices for cultivation and acquisition of soil and weather data will facilitate comparisons and allow for inclusion of specific environmental variables in quantitative models. Such an effort has recently begun with the Genomes to Fields (G2F) Initiative, which aims to characterize genotype by environment effects by growing inbred and hybrid lines of maize in 22 environments, with standardized weather and soil data collection in each environment [47]. Ecophysiological models and evolutionary ecology are becoming useful components of quantitative genetic analyses seeking to describe G x E. QTL-based ecophysiological models can specifically model components of the environment and predict the outcome of a given genotype-environment pair [48].

Quantitative studies of the ionome can more completely characterize the genes controlling the ionome, contextualize genes within networks, and link genetic networks to adaptive response if experiments are conducted in a broader range of species, environments, and tissues and merged with other –omics data. To fully capture genetic regulation of the ionome, it will be necessary to view the ionome as a network and advance the use of multivariate analysis in quantitative studies. PCA and LDA have been shown to separate out groups of mutants and distinguish plants grown in different environments based on the ionome as a whole. Nutrient balances, isometric log ratios of elements and groups of elements, have also been proposed as a method of multivariate ionomic analysis [49]. Integration of techniques such as transcriptomics

or metabolomics with ionomics will further gene and gene network-level understanding of the ionome [50, 51]. Progress in next-generation phenotyping will aid in linking the ionome to adaptive traits [52]. Visualization techniques involving X-ray fluorescence, X-ray absorption spectrometry, and mass spectrometry are being developed to image the cellular and subcellular localization of elements and trace the movement of elements [53, 54]. A comprehensive view of ionomic regulation will improve as studies in the field include more data types and various environmental conditions.

OUTLINE OF THE DISSERTATION

This work advances the understanding of mineral nutrient regulation in the crop species maize (Zea mays L.) through a series of experiments utilizing quantitative genetics, multivariate approaches, and gene expression analysis. Maize is both a model plant species, with extensive genetic resources [55], and a global staple crop, with several practical applications for human nutrition and energy production [56]. Maize genetic diversity exceeds that of any other model organism [57]. This diversity has been cultivated over thousands of years and variable environments, making it ideal for the study of the genetic by environment interactions that determine the ionome. Although variation in element homeostasis is expected across different organisms, comparative genomics has been successfully used in previous metabolomics and ionomics studies [58, 59], suggesting that findings in maize can be extended to other plant species.

The first chapter of this thesis describes a QTL by environment analysis of kernel element content in the maize intermated B73 x Mo17 recombinant inbred (IBM) population [20], a population particularly suited for quantitative genetic analyses as its high level of intermating

and genetic recombination allows for improved mapping resolution [60]. This analysis was carried out on element profiles from seeds of the IBM population grown in 10 different field environments to map loci contributing to element accumulation in the seed. Varying field environments also allowed for detection of loci exhibiting interactions with environment.

The second chapter details QTL analysis using a multivariate technique. This extension beyond single-element QTL mapping was motivated by previous descriptions of the ionome as an extensively correlated network [6]. This approach isolated locations of the genome contributing to variation in multiple elements through QTL mapping on multi-element traits derived from PCA. Multi-element traits served as a means to approach the ionome as an integrated web of elements and find genetic regulators shared between different elements. These results suggest that single-element and multi-element techniques should be used as complimentary methods to maximize detection of genetic loci contributing to seed element accumulation.

The third and final chapter describes a gene expression study using RNA collected from roots of the IBM population grown in greenhouse conditions. Gene expression in the root has been repeatedly shown to impact the ionome of seeds and leaves [2, 30, 61]. This work identified gene expression networks in the maize root using co-expression and expression QTL (eQTL) analyses. Genetic networks that act in the root and potentially impact the leaf and/or seed ionome were characterized by relating the root-based results with previously mapped loci for leaf and seed elemental profiles. Candidate genes were identified for known ionome QTL and QTL were contextualized within broader genetic networks.

By employing an integrative, multi-staged analysis with sets of ionomic and gene expression data across various environments, these experiments have identified genetic loci and

regulatory networks in maize underlying element accumulation. Further exploration of these candidate genes and regulatory mechanisms can inform on genetic control of adaptive traits and provide foundational knowledge for selective breeding of crops that efficiently produce fuel and nutrients in increasingly variable environments.

REFERENCES

1. Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE et al. Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana. Nat Biotechnol. 2003;21: 1215-1221.
2. Baxter I, Hosmani PS, Rus A, Lahner B, Borevitz JO, Muthukumar B et al. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. PLoS Genet. 2009;5: e1000492.
3. Baxter IR, Vitek O, Lahner B, Muthukumar B, Borghi M, Morrissey J et al. The leaf ionome as a multivariable system to detect a plant's physiological status. Proceedings of the National Academy of Sciences. 2008;105: 12081-12086.
4. Chen Z, Watanabe T, Shinano T, Ezawa T, Wasaki J, Kimura K et al. Element interconnections in Lotus japonicus: A systematic study of the effects of element additions on different natural variants. Soil science and plant nutrition. 2009;55: 91-101.
5. Broadley MR, Hammond JP, King GJ, Astley D, Bowen HC, Meacham MC et al. Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in Brassica oleracea. Plant Physiol. 2008;146: 1707-1720.
6. Baxter I. Ionomics: studying the social network of mineral nutrients. Current Opinion in Plant Biology. 2009;12: 381-386.
7. Chen Z, Watanabe T, Shinano T, Osaki M. Rapid characterization of plant mutants with an altered ion-profile: a case study using Lotus japonicus. New Phytol. 2009;181: 795-801.
8. Ziegler G, Terauchi A, Becker A, Armstrong P, Hudson K, Baxter I. Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. The Plant Genome. 2013;6:
9. Kamiya T, Borghi M, Wang P, Danku JMC, Kalmbach L, Hosmani PS et al. The MYB36 transcription factor orchestrates Casparian strip formation. Proceedings of the National Academy of Sciences. 2015;112: 10533-10538.
10. Chao DY, Gable K, Chen M, Baxter I, Dietrich CR, Cahoon EB et al. Sphingolipids in the Root Play an Important Role in Regulating the Leaf Ionome in Arabidopsis thaliana. Plant Cell. 2011;23: 1061-1081.
11. Rus A, Baxter I, Muthukumar B, Gustin J, Lahner B, Yakubova E et al. Natural variants of At HKT1 enhance Na+ accumulation in two wild populations of Arabidopsis. PLoS Genet. 2006;2: e210.
12. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J et al. Variation in Molybdenum Content Across Broadly Distributed Populations of *Arabidopsis thaliana* Is Controlled by a Mitochondrial Molybdenum Transporter (*MOT1*). PLoS Genet. 2008;4.

13. Poormohammad Kiani S, Trontin C, Andreatta M, Simon M, Robert T, Salt DE et al. Allelic Heterogeneity and Trade-Off Shape Natural Variation for Response to Soil Micronutrient. PLOS Genetics. 2012;8: e1002814.
14. Morrissey J, Baxter IR, Lee J, Li L, Lahner B, Grotz N et al. The ferroportin metal efflux proteins function in iron and cobalt homeostasis in Arabidopsis. The Plant Cell Online. 2009;21: 3326-3338.
15. Loudet O, Saliba-Colombani V, Camilleri C, Calenge F, Gaudon V, Koprivova A et al. Natural variation for sulfate content in Arabidopsis thaliana is highly controlled by APR2. Nature genetics. 2007;39: 896-900.
16. Koprivova A, Harper AL, Trick M, Bancroft I, Kopriva S. Dissection of the control of anion homeostasis by associative transcriptomics in Brassica napus. Plant physiology. 2014;166: 442-450.
17. Chao D-Y, Silva A, Baxter I, Huang YS, Nordborg M, Danku J et al. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in Arabidopsis thaliana. PLoS Genet. 2012;8: e1002923.
18. Chao D-Y, Chen Y, Chen J, Shi S, Chen Z, Wang C et al. Genome-wide association mapping identifies a new arsenate reductase enzyme critical for limiting arsenic accumulation in plants. PLoS Biol. 2014;12: e1002009.
19. Baxter IR, Gustin JL, Settles AM, Hoekenga OA. Ionomic characterization of maize kernels in the intermated B73 x Mo17 population. Crop Science. 2013;53: 208-220.
20. Asaro A, Ziegler G, Ziyomo C, Hoekenga OA, Dilkes BP, Baxter I. The Interaction of Genotype and Environment Determines Variation in the Maize Kernel Ionome. G3: Genes Genomes Genetics. 2016;6: 4175-4183.
21. 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: 139-153.
22. Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, Connolly EL et al. Integration of experiments across diverse environments identifies the genetic determinants of variation in Sorghum bicolor seed element composition. Plant physiology. 2016pp-01971.
23. Ziegler GR. Development of tools to enable high-throughput elemental analysis and their application to soybean mutant identification and genome wide association studies. 2014
24. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant methods. 2013;9: 1.
25. Borghi M, Rus A, Salt DE. Loss-of-function of Constitutive Expresser of Pathogenesis Related Genes5 affects potassium homeostasis in Arabidopsis thaliana. PLoS One. 2011;6: e26360.
26. Tian H, Baxter IR, Lahner B, Reinders A, Salt DE, Ward JM. Arabidopsis NPCC6/NaKR1 is a phloem mobile metal binding protein necessary for phloem function and root meristem maintenance. Plant Cell. 2010;22: 3963-3979.
27. Baxter I, Brazelton JN, Yu D, Huang YS, Lahner B, Yakubova E et al. A Coastal Cline in Sodium Accumulation in *Arabidopsis thaliana* Is Driven by Natural Variation of the Sodium Transporter AtHKT1;1. PLoS Genet. 2010;6: e1001193.
28. Sanchez-Rodriguez E, del Mar Rubio-Wilhelmi M, Cervilla LM, Blasco B, Rios JJ, Leyva R et al. Study of the ionome and uptake fluxes in cherry tomato plants under moderate water stress conditions. Plant and soil. 2010;335: 339-347.
29. Zhang M, Pinson SR, Tarpley L, Huang XY, Lahner B, Yakubova E et al. Mapping and

validation of quantitative trait loci associated with concentrations of 16 elements in unmilled rice grain. Theor Appl Genet. 2014;127: 137-165.
30. Ghandilyan A, Ilk N, Hanhart C, Mbengue M, Barboza L, Schat H et al. A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three Arabidopsis thaliana RIL populations. J Exp Bot. 2009;60: 1409-1425.
31. Buescher E, Achberger T, Amusan I, Giannini A, Ochsenfeld C, Rus A et al. Natural genetic variation in selected populations of Arabidopsis thaliana is associated with ionomic differences. PLoS One. 2010;5: e11081.
32. Xu Y-F, An D-G, Liu D-C, Zhang A-M, Xu H-X, Li B. Mapping QTLs with epistatic effects and QTL x treatment interactions for salt tolerance at seedling stage of wheat. Euphytica. 2012;186: 233-245.
33. Baxter I, Dilkes BP. Elemental Profiles Reflect Plant Adaptations to the Environment. Science. 2012;336: 1661-1663.
34. Pineau C, Loubet S, Lefoulon C, Chalies C, Fizames C, Lacombe B et al. Natural variation at the FRD3 MATE transporter locus reveals cross-talk between Fe homeostasis and Zn tolerance in Arabidopsis thaliana. PLoS Genet. 2012;8: e1003120.
35. Arrivault S, Senger T, Kramer U. The Arabidopsis metal tolerance protein AtMTP3 maintains metal homeostasis by mediating Zn exclusion from the shoot under Fe deficiency and Zn oversupply. The Plant Journal. 2006;46: 861-879.
36. Ray DK, Mueller ND, West PC, Foley JA. Yield trends are insufficient to double global crop production by 2050. PloS one. 2013;8: e66428.
37. Bohra A, Sahrawat KL, Kumar S, Joshi R, Parihar AK, Singh U et al. Genetics-and genomics-based interventions for nutritional enhancement of grain legume crops: status and outlook. Journal of applied genetics. 2015;56: 151-161.
38. Gangashetty PI, Motagi BN, Pavan R, Roodagi MB. Breeding Crop Plants for Improved Human Nutrition Through Biofortification: Progress and Prospects. Springer; 2016. p. 35-76.
39. Lopez-Arredondo DL, Leyva-Gonzalez MA, Gonzalez-Morales SI, Lopez-Bucio J, Herrera-Estrella L. Phosphate nutrition: improving low-phosphate tolerance in crops. Annual Review of Plant Biology. 2014;65: 95-123.
40. Li Z, Gao Q, Liu Y, He C, Zhang X, Zhang J. Overexpression of transcription factor ZmPTF1 improves low phosphate tolerance of maize by regulating carbon metabolism and root growth. Planta. 2011;233: 1129-1143.
41. Yi K, Wu Z, Zhou J, Du L, Guo L, Wu Y et al. OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. Plant physiology. 2005;138: 2087-2096.
42. Dai X, Wang Y, Yang A, Zhang W-H. OsMYB2P-1, an R2R3 MYB transcription factor, is involved in the regulation of phosphate-starvation responses and root architecture in rice. Plant Physiology. 2012;159: 169-183.
43. Korshunova YO, Eide D, Clark WG, Guerinot ML, Pakrasi HB. The IRT1 protein from Arabidopsis thaliana is a metal transporter with a broad substrate range. Plant molecular biology. 1999;40: 37-44.
44. Ghandilyan A, Barboza L, Tisné S, Grainer C, Reymond M, Koorneef M et al. Genetics analysis identifies quantitative trait loci controlling rosette mineral concentrations in *Arabidopsis thaliana* under drought. New Phytol. 2009;184: 180-192.

45. Vikram P, Swamy BPM, Dixit S, Ahmed H, Cruz MTS, Singh AK et al. Bulk segregant analysis: "An effective approach for mapping consistent-effect drought grain yield QTLs in rice". Field Crops Research. 2012;134: 185-192.
46. Quadir QF, Watanabe T, Chen Z, Osaki M, Shinano T. Ionomic response of Lotus japonicus to different root-zone temperatures. Soil science and plant nutrition. 2011;57: 221-232.
47. Springer, NM. Genomes to Fields: Translating Our Understanding of the Genome to Predictions of Performance in the Field. 2016; Plant and Animal Genome; 2016.
48. El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. Genotype x environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science. 2014;19: 390-398.
49. Parent SE, Parent LE, Rozane DE, Natale W. Plant ionome diagnosis using sound balances: case study with mango (Mangifera Indica). Front Plant Sci. 2013;4: 449.
50. Huang X-Y, Salt DE. Plant ionomics: From elemental profiling to environmental adaptation. Molecular plant. 2016;9: 787-797.
51. Kumar A, Pathak RK, Gupta SM, Gaur VS, Pandey D. Systems Biology for Smart Crops and Agricultural Innovation: Filling the Gaps between Genotype and Phenotype for Complex Traits Linked with Robust Agricultural Productivity and Sustainability. Omics: a journal of integrative biology. 2015;19: 581-601.
52. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet. 2013;126: 867-887.
53. Zhao F-J, Moore KL, Lombi E, Zhu Y-G. Imaging element distribution and speciation in plant cells. Trends in plant science. 2014;19: 183-192.
54. Wang YX, Specht A, Horst WJ. Stable isotope labelling and zinc distribution in grains studied by laser ablation ICP-MS in an ear culture system reveals zinc transport barriers during grain filling in wheat. New Phytologist. 2011;189: 428-437.
55. Wallace JG, Larsson SJ, Buckler ES. Entering the second century of maize quantitative genetics. Heredity. 2014;112: 30-38.
56. Moshelion M, Altman A. Current challenges and future perspectives of plant and agricultural biotechnology. Trends in biotechnology. 2015;33: 337-342.
57. Buckler ES, Gaut BS, McMullen MD. Molecular and functional diversity of maize. Current opinion in plant biology. 2006;9: 172-176.
58. Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD. Quantitative Trait Loci and Comparative Genomics of Cereal Cell Wall Composition. Plant Physiology. 2003;132: 263-271.
59. Sankaran RP, Huguet T, Grusak MA. Identification of QTL affecting seed mineral concentrations and content in the model legume Medicago truncatula. Theor Appl Genet. 2009;119: 241-253.
60. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D et al. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant molecular biology. 2002;48: 453-461.
61. Duan G, Kamiya T, Ishikawa S, Arao T, Fujiwara T. Expressing ScACR3 in rice enhanced arsenite efflux and reduced arsenic accumulation in rice grains. Plant and Cell Physiology. 2012;53: 154-163.

CHAPTER 2:

THE INTERACTION OF GENOTYPE AND ENVIRONMENT DETERMINES VARIATION IN THE MAIZE KERNEL IONOME

Alexandra Asaro, Greg Ziegler, Cathrine Ziyomo, Owen A. Hoekenga,

Brian P. Dilkes, Ivan Baxter

This chapter consists of work that was peer reviewed and published in G3: Genes Genomes Genetics (2016). AA conducted all data analysis. AA interpreted results, created figures, and wrote the manuscript with assistance from BD and IB. GZ and CZ collected ionomics data and helped with statistical analysis.

ABSTRACT

Plants obtain soil-resident elements that support growth and metabolism from the water-flow facilitated by transpiration and active transport processes. The availability of elements in the environment interacts with the genetic capacity of organisms to modulate element uptake through plastic adaptive responses, such as homeostasis. These interactions should cause the elemental contents of plants to vary such that the effects of genetic polymorphisms will be dramatically dependent on the environment in which the plant is grown. To investigate genotype by environment interactions underlying elemental accumulation, we analyzed levels of elements in maize kernels of the Intermated B73 x Mo17 (IBM) recombinant inbred population grown in 10 different environments spanning a total of six locations and five different years. In analyses conducted separately for each environment, we identified a total of 79 quantitative trait loci controlling seed elemental accumulation. While a set of these QTL were found in multiple environments, the majority were specific to a single environment, suggesting the presence of genetic by environment interactions. To specifically identify and quantify QTL by environment interactions (QEIs), we implemented two methods: linear modeling with environmental covariates and QTL analysis on trait differences between growouts. With these approaches, we found several instances of QEI, indicating that elemental profiles are highly heritable, interrelated, and responsive to the environment.

INTRODUCTION

The intake, transport, and storage of elements are key processes underlying plant growth and survival. A plant must balance mineral levels to prevent accumulation of toxic concentrations of elements while taking up essential elements for growth. Food crops must strike similar balances to provide healthy nutrient contents of edible tissues. Adaptation to variation in soil, water, and temperature requires that plant genomes encode flexible regulation of mineral physiology to achieve homeostasis [1]. This regulation must be responsive to both the availability of each regulated element in the environment and the levels of these elements at the sites of use within the plant. Understanding how the genome encodes responses to element limitation or toxic excess in nutrient-poor or contaminated soils will help to achieve targeted crop improvements and sustain our rapidly growing human population [2].

The concentrations of elements in a plant sample provide a useful read-out for the environmental, genetic and physiological processes important for plant adaptation. We and others developed high-throughput and inexpensive pipelines to detect and quantitate 20 different elemental concentrations by inductively coupled plasma mass spectrometry (ICP-MS). This process, termed ionomics, is the quantitative study of the complete set of mineral nutrients and trace elements in an organism (its ionome) [3]. In crop plants such as maize and soybean, seed element profiles make an ideal study tissue as seeds provide a read-out of physiological status of the plant and are the food source.

Quantitative genetics using structured recombinant inbred populations is a powerful tool for dissecting the factors underlying elemental accumulation and relationships. By breaking up linkage blocks through recombination and then fixing these new haplotypes of diverse loci into mosaic sets of lines, these populations allow similar sets of alleles to be repeatedly tested in

diverse environments [4]. A variety of quantitative statistical approaches can then be used to identify QTL by environment interactions (QEI).

Here, we used elemental profiling of a maize recombinant inbred population grown in multiple environments to identify QTL and QEI underlying elemental accumulation. We sought both environmental and genetic determinants by implementing single-environment QTL mapping and analyses of combined data from multiple environments. Overall, we detected 79 loci controlling elemental accumulation, many of which were environment-specific, and identified loci exhibiting significant QEI.

MATERIALS & METHODS

Field Growth and Data Collection

Population and field growth. Subsets of the Intermated B73 x Mo17 (IBM) recombinant inbred population were grown in 10 different environments: Homestead, Florida in 2005 (220 lines) and 2006 (118 lines), West Lafayette, Indiana in 2009 (193 lines) and 2010 (168 lines), Clayton, North Carolina in 2006 (197 lines), Poplar Ridge, New York in 2005 (256 lines), 2006 (82 lines), and 2012 (168 lines), Columbia, Missouri in 2006 (97 lines), and Limpopo, South Africa in 2010 (87 lines). In all but three environments, NY05, NC06, and MO06, one replicate was sampled per line. In NY05, 3 replicates of 199 lines, 2 replicates of 50 lines, and 1 replicate of 7 lines were sampled. A replicate is considered pooled ears from a row. Several ears were harvested and kernels were subsampled from pooled ears from the row. After harvesting, seeds were stored in local temperature and humidity controlled seed storage rooms. Subsequently they were shipped to the ionomics lab where they were stored in temperature-controlled conditions. Because each batch of seed was treated identically, any losses in weight or increases

in weight due to differing hydration should not affect the relative, weight-adjusted concentrations used for analysis. We do not expect any changes in ion composition due to storage. Table S1 includes planting dates and line numbers after outlier removal and genotype matching. After outlier removal, 199 of the 233 unique lines in the experiment were present in 3 or more of the 10 environments. 106 lines were present in 7 or more of the environments.

Elemental Profile Analysis

Elemental profile analysis is conducted as a standardized pipeline in the Baxter Lab. The methods used for elemental profile analysis are as described in Ziegler et al. [5]. Descriptions taken directly are denoted by quotation marks.

Sample preparation and digestion. Lines from the IBM population from each environment were analyzed for the concentrations of 20 elements. "Seeds were sorted into 48-well tissue culture plates, one seed per well. A weight for each individual seed was determined using a custom built weighing robot. The weighing robot holds six 48-well plates and maneuvers each well of the plates over a hole which opens onto a 3-place balance. After recording the weight, each seed was deposited using pressurized air into a 16×110 mm borosilicate glass test tube for digestion. The weighing robot can automatically weigh 288 seeds in approximately 1.5 hours with little user intervention."

"Seeds were digested in 2.5 mL concentrated nitric acid (AR Select Grade, VWR) with internal standard added (20 ppb In, BDH Aristar Plus). Seeds were soaked at room temperature overnight, then heated to 105°C for two hours. After cooling, the samples were diluted to 10 mL using ultrapure 18.2 MΩ water (UPW) from a Milli-Q system (Millipore). Samples were stirred with a custom-built stirring rod assembly, which uses plastic stirring rods to stir 60 test tubes at a time. Between uses, the stirring rod assembly was soaked in a 10% HNO_3 solution. A second

dilution of 0.9 mL of the 1st dilution and 4.1 mL UPW was prepared in a second set of test tubes. After stirring, 1.2 mL of the second dilution was loaded into 96 well autosampler trays."

Ion Coupled plasma mass spectrometry analysis. Elemental concentrations of B, Na, Mg, Al, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, and Cd "were measured using an Elan 6000 DRC-e mass spectrometer (Perkin-Elmer SCIEX) connected to a PFA microflow nebulizer (Elemental Scientific) and Apex HF desolvator (Elemental Scientific). Samples were introduced using a SC-FAST sample introduction system and SC4-DX autosampler (Elemental Scientific) that holds six 96-well trays (576 samples)." Measurements were taken with dynamic reaction cell (DRC) collision mode off. "Before each run, the lens voltage and nebulizer gas flow rate of the ICP-MS were optimized for maximum Indium signal intensity (>25,000 counts per second) while also maintaining low CeO+/Ce+ (<0.008) and Ba++/Ba+ (<0.1) ratios. This ensures a strong signal while also reducing the interferences caused by polyatomic and double-charged species. Before each run a calibration curve was obtained by analyzing six dilutions of a multi-element stock solution made from a mixture of single-element stock standards (Ultra Scientific). In addition, to correct for machine drift both during a single run and between runs, a control solution was run every tenth sample. The control solution is a bulk mixture of the remaining sample from the second dilution. Using bulked samples ensured that our controls were perfectly matrix matched and contained the same elemental concentrations as our samples, so that any drift due to the sample matrix would be reflected in drift in our controls. The same control mixture was used for every ICP-MS run in the project so that run-to-run variation could be corrected. A run of 576 samples took approximately 33 hours with no user intervention. The time required for cleaning of the instrument and sample tubes as well as the digestions and transfers necessary to set up the run limit the throughput to three 576 sample runs per week."

Computational Analysis

Drift correction and analytical outlier removal. Analytical outliers were removed from single-seed measurements using a method described by Davies and Gather [6]. Briefly, values were considered an outlier and removed from further analysis if the median absolute deviation (MAD), calculated based on the line and location where the seed was grown, was greater than 6.2.

Normalization for seed weight by simply dividing each seed's solution concentration by sample weight resulted in a bias where smaller seeds often exhibited a higher apparent elemental concentration, especially for elements whose concentration is at or near the method detection limit. This bias is likely either a result of contamination during sample processing, a systematic over or under reporting of elemental concentrations by the ICP-MS, or a violation of the underlying assumption that elemental concentration in seeds scales linearly with seed weight. Instead, we developed a method taking residuals from the following linear model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

where Y is the non-weight normalized measure of elemental concentration for each seed after digestion, β_0 is the population mean, X_1 is the seed weight, X_2 is the analytical experiment the seed was run in (to further correct for run-to-run variation between analytical experiments), and e is the residual (error) term. The residuals in this linear model represent how far each data point departs from our assumption that analyte concentration will scale linearly with sample weight. If all samples have the same analyte concentration then the linear model will be able to perfectly predict analyte concentration from weight and the residuals will all equal zero. However, if a sample has a higher or lower concentration of an analyte then the general population being measured, then it will have a residual whose value represents the estimated concentration

difference from the population mean. For this reason, we have termed this value the estimated concentration difference from the mean (ECDM).

Heritability calculation. Broad-sense heritability was calculated for seed weight and 20 elements across environments and within three environments for which we had substantial replicate data. To estimate the broad-sense heritability across 10 environments, the total phenotypic variance was partitioned into genetic and environmental variance, with the broad-sense heritability being the fraction of phenotypic variance that is genetic. This was done using an unbalanced, type II analysis of variance (ANOVA) in order to account for the unbalanced common line combinations across environments. Two models were fit using the *lmfit* function in R. The first model included genetic variance as the first term and environmental variance as the second. The second model had the opposite form. The variances for genetic or environmental components were obtained using the *anova* function on the model in which that component was the second term. Broad-sense heritability was calculated by dividing the genetic variance by the total (genetic plus environmental plus residuals) variance. Heritability was calculated within environments for NY05, NC06, and MO06. Data with outliers designated as NA was used for each environment. For each element within an environment, lines with NA were removed and lines with only 1 replicate were removed, leaving only lines with 2 or more replicates. The heritability was then calculated for seed weight and each element using the *lmfit* and *anova* functions to obtain the variances for the genetic component and the residuals. Broad-sense heritability was calculated as the proportion of total variance (genetic plus residuals) explained by the genetic component.

QTL mapping: elemental traits. The R package R/qtl was used for QTL mapping. For each of the 10 environments, elemental trait line averages and genotypes for all lines, 4,217

biallelic single nucleotide polymorphisms (SNPs) distributed across all 10 maize chromosomes, were formatted into an R/qtl cross object. The *stepwiseqtl* function was used to implement the stepwise method of QTL model selection for 21 phenotypes (seed weight, 20 elements). The max number of QTL allowed for each trait was set at 10 and the penalty for addition of QTL was set as the 95th percentile LOD score from 1000 *scanone* permutations, with imputation as the selected model for *scanone*. A solely additive model was used; epistatic and interaction effects were not considered and thus heavy and light interaction penalties were set at 0. QTL positions were optimized using *refineqtl*, which considers each QTL one at a time, in random order, iteratively scanning in order to move the QTL to the highest likelihood position. QTL models for each trait in each environment were obtained using this procedure. QTL within 5 cM of each other were designated as the same QTL.

QTL by environment analysis: linear model comparison. Linear modeling was used to determine instances and strength of QEI using all data from two years within three locations (FL, IN, NY). The specific growouts analyzed together were FL05, FL06, IN09, IN10, NY05, and NY12. FL, IN, and NY were then used as covariates in QTL analysis. Two QTL models, one with location as an additive and interactive covariate and one with location as only an additive covariate, were fit for each phenotype (sample weight, 20 elements) using the *scanone* function in R/qtl,

$$y_i = \mu + \beta_g g_i + \beta_x x_i + \gamma g_i x_i + \varepsilon_i \qquad (1)$$

$$y_i = \mu + \beta_g g_i + \beta_x x_i + \varepsilon_i \qquad (2)$$

where y_i is the phenotype of individual i, g_i is the genotype of individual i, and x_i is the location of individual i. B_g and B_x are additive effects of genotype and environment, respectively, and γ is the effect of genotype by environmental interaction. LOD scores for each marker using model

(2) were subtracted from LOD scores for each marker using model (1) to the isolate genetic by location effect. QTL by location interaction was determined as QTL with a significant LOD score after subtraction. The significance threshold was calculated from 1000 permutations of the three step procedure (fitting the two models and then subtracting LOD scores) and taking the 95th percentile of the highest LOD score.

QTL by environment analysis: mapping on within-location differences. QTL were mapped on phenotypic differences between common lines grown over two years at a single location. This procedure was used to compare FL05 and FL06, IN09 and IN10, and NY05 and NY12 by calculating the differences for each trait value between common lines in location pairs (FL05-FL06, IN09-IN10, NY05-NY12) and using these differences for analysis using the previously described *stepwiseqtl* mapping and permutation procedure.

Data Availability

All data and scripts are available on Ionomics Hub (iHUB) in the Maize Database at www.ionomicshub.org.

RESULTS

Genetic Regulation of Elemental Traits

The data used for this study is comprised of 20 elements measured in the seeds from the *Zea mays* L. Intermated B73 x Mo17 recombinant inbred line (IBM) population grown in 10 different location/year settings. The IBM population is a widely studied maize population of 302 intermated recombinant inbred lines, each of which have been genotyped with a set of 4,217 bi-allelic single nucleotide polymorphism (SNP) genetic markers [7]. The four rounds of

intermating and subsequent inbreeding generated increased recombination and a longer genetic map for the IBM than for typical biparental recombinant inbred line populations. The number of individuals, marker density, and greater recombination facilitates more precise QTL localization than a standard RIL population [8–13]. This greater resolution reduces the number of genes within a QTL support interval, increasing the utility of QTL mapping as a hypothesis test for shared genetic regulation of multiple traits and promoting discovery of the molecular identity of genes affecting QTL. For this study, subsets of the IBM population were grown at Homestead, Florida in 2005 (FL05) and 2006 (FL06), West Lafayette, Indiana in 2009 (IN09) and 2010 (IN10), Clayton, North Carolina in 2006 (NC06), Poplar Ridge, New York in 2005 (NY05), 2006 (NY06), and 2012 (NY12), Columbia, Missouri in 2006 (MO06), and Limpopo, South Africa in 2010 (SA10) (Table S1). While very few of the 233 unique IBM lines in the experiment were grown in all environments, 106 of the 233 lines were grown in 7 or more environments and 199 were grown in 3 or more environments. Within each growout, all samples were treated identically: seeds from all environments were stored in temperature and humidity controlled storage rooms after harvest and then shipped to the ionomics lab. We do not expect any change in ion composition from storage within a growout, however we cannot rule out that some of the differences between growouts might be due to slightly different moisture content. These differences are not likely to account for the genetic by environment interactions we observe as they should have similar effects on all lines. Single seeds were profiled for the quantities of 20 elements using ICP-MS. These measurements were normalized to seed weight and technical sources of variation using a linear model, with the resulting values used as the elemental traits for all analyses [14]. After outlier removal, seed element phenotypes were

derived by averaging line replicates (kernels subsampled out of pooled ears from a row) within an environment.

Variation in the elemental traits was affected by both environment and genotype. Elemental traits generally exhibited lower heritability among genotypes grown across multiple environments than among genotype replicates within a single environment (Table 1). The broad-sense heritability (H^2) of seed weight, 15 of 21 elements in NY05, 13 of 21 elements in NC06, and 13 of 21 elements in MO06 exceeded 0.60. Elements exhibiting low heritability within environments corresponded to the elements that are prone to analytical artifacts or present near the limits of detection by our methods, such as B, Al, and As. Seven elements had a broad sense heritability of at least 0.45 in a single environment (NY05, NC06, and NY06) but less than 0.1 across all environments. This decrease in heritability across the experiment, which was particularly striking for Mg, P, S, and Ni, is consistent with strong genotype by environment interactions governing the accumulation of these elements.

Table 1. Broad-sense Heritability (H^2) of Element Concentrations.

Trait	All env	NY05	NC06	MO06
Seed Weight	0.30	0.59	0.69	0.89
B	0.02	0.35	0.51	0.06
Na	0.07	0.34	0.23	0.19
Mg	0.04	0.77	0.69	0.75
Al	0.07	0.39	0.50	0.08
P	0.03	0.62	0.69	0.33
S	0.05	0.73	0.77	0.51
K	0.06	0.69	0.72	0.36
Ca	0.12	0.65	0.63	0.77
Mn	0.14	0.80	0.80	0.75
Fe	0.07	0.76	0.73	0.63
Co	0.06	0.65	0.54	0.42
Ni	0.05	0.84	0.54	0.82

Cu	0.17	0.80	0.75	0.92
Zn	0.07	0.68	0.73	0.86
As	0.02	0.37	0.45	0.01
Se	0.03	0.32	0.35	0.68
Rb	0.03	0.49	0.45	0.69
Sr	0.06	0.61	0.48	0.53
Mo	0.23	0.85	0.73	0.96
Cd	0.36	0.71	0.69	0.24

All env: Line replicate averages from each location
NY05: 50 lines with 2 reps, 199 lines with 3 reps
NC06: 121 lines with 2 reps, 53 lines with 3 reps, 4 lines with 4 reps
MO06: 50 lines with 2 reps, 18 lines with 3 reps
*outliers for each element calculated with outlier removal function, designated as NA
*for each single environment, for each trait, only lines w/o missing data and with reps >1 used to calculate heritability

A stepwise algorithm, implemented via *stepwiseqtl* in the R package R/qtl [15], was used to map QTL for seed weight and 20 seed elemental phenotypes. The stepwise algorithm iterates through the genome and tests for significant allelic effects of each marker on a phenotype. Forward and backward regression generates the final genome-wide QTL models for each trait. This QTL mapping procedure on 21 traits was completed as a separate analysis for each subset of lines from the IBM populations grown in each of the 10 environments. For the sake of completeness and to comprehensively investigate all of the traits we had access to, all elemental traits in each environment were tested, even in cases where heritability for a given element was low in an environment. QTL significance were determined using the 95^{th} percentile threshold from 1000 *scanone* permutations as a penalty score for adding QTL to the stepwise model [16]. We examined the relationship between the heritability of an element in a given environment and number of QTL identified in that environment (Fig S1). As expected, elements with very low heritability had few to no QTLs while larger numbers of QTLs were identified for higher heritability elements.

The environmental dependence on QTL detection was first estimated by identifying QTL common to multiple environments. If QTL detected in two or more growouts affected the same element and localized within 5 cM of each other, they were considered to be the same locus. Across the 10 environments, a total of 79 QTL were identified for seed weight and 18 of the 20 elemental traits tested (none for Al or Co) (Fig 1B &C). Of these QTL, 63 were detected in a single environment and 16 were detected in multiple environments. The 16 QTL found in multiple environments included QTL detected in nearly all of the environments and QTL detected in only two. One QTL for Mo accumulation, on chromosome 1 in the genetic region containing the maize ortholog of the Arabidopsis molybdenum transporter MOT1 [17], was found in nine environments (Fig 1A). Another QTL affecting Cd accumulation, on chromosome 2 and without a clear candidate gene, was found in eight environments. Other QTL were only present in a smaller set of environments, such as the QTL for Ni accumulation on chromosome 9, which was found in five environments (Fig 1D). The strength of association and percent variance explained showed strong differences between environments even for these QTL that were detected in multiple environments (Table S2).

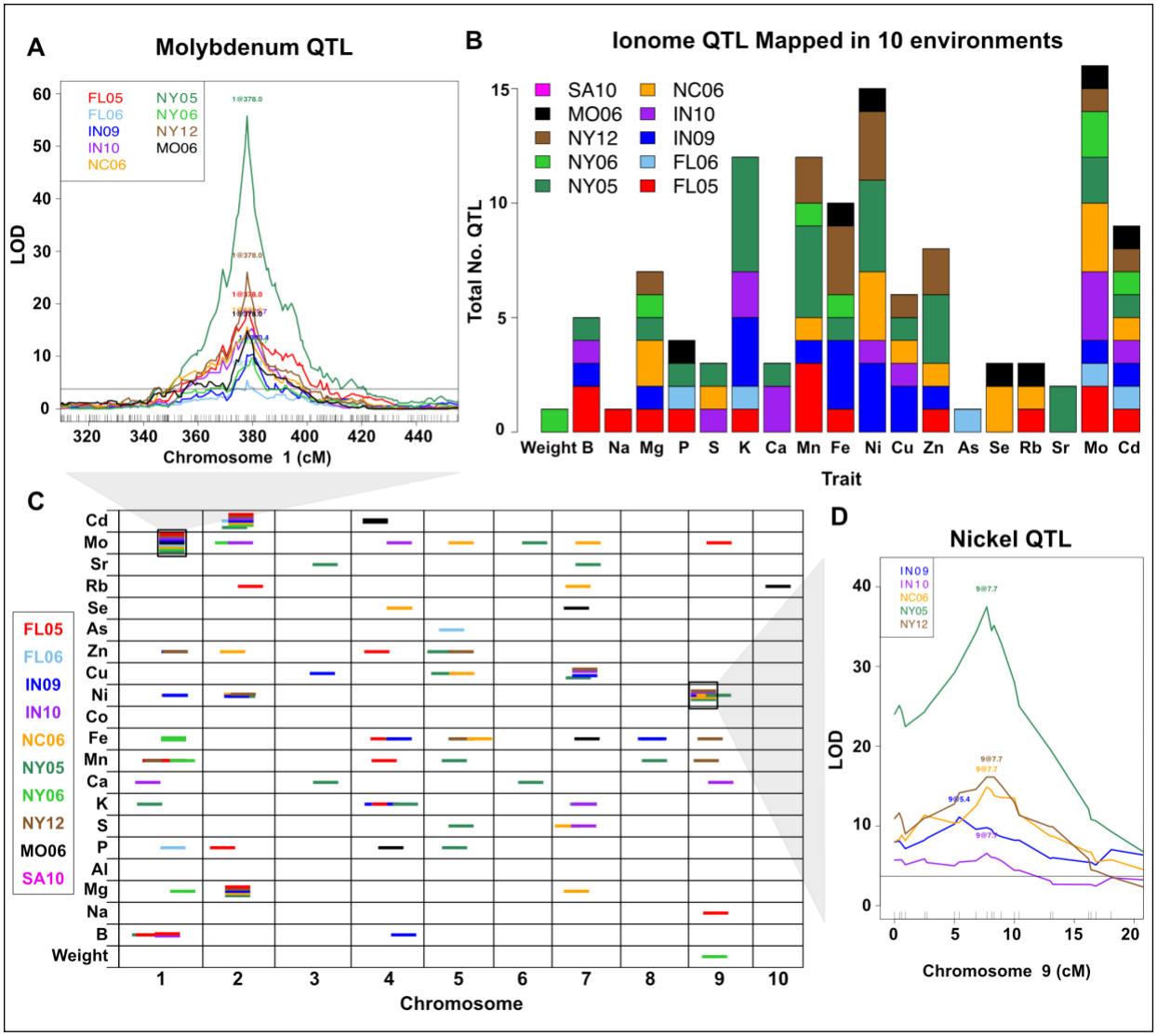


Fig 1. Ionome QTL from 10 Environments. QTL identified for seed weight and 20 element accumulation traits using the IBM RIL population grown in 10 environments. (A) QTL on chromosome 1 affecting variation in molybdenum accumulation. An interval of Chr1 is shown on the x-axis in centi-Morgans (cM). The LOD score for the trait-genotype association is shown on the y-axis. The horizontal line is a significance threshold from 1000 random permutations (α= 0.05). The LOD profiles are plotted for all environments in which the highlighted QTL was detected. (B) Total number of QTL detected for each trait, colored by environment. (C) Significant QTL (α= 0.05) for each trait. QTL location is shown across the 10 maize chromosomes (in cM) on the x-axis. Dashes indicate QTL, with environment in which QTL was found designated by color. All dashes are the same length for visibility. The two black boxes around dashes correspond to LOD profiles traces in (A) and (D). (D) Stepwise QTL mapping output for nickel on chromosome 9. LOD profiles are plotted for all environments in which the QTL is significant.

As seen in the full-genome view of all QTL colored by environment (Fig 1C), there is a high incidence of QTL found in single locations. There are three hypotheses that could explain the large proportion of QTL found only in a single location: 1) strong QTL by environment interaction effects, 2) false positive detection of a QTL in an individual location and 3) false negative assessment of QTL absence due to genetic action but statistical assessment below the permutation threshold in other environments. To reduce the risk of false positives in a single environment's QTL set, the significance threshold was raised to the 99th percentile, where 31 of the 63 environment-specific QTL remained significant. Despite the large number of trait/environment combinations tested (20 traits in 10 environments), the number of QTLs detected is much larger than the null expectation derived from a Bonferroni correction: 10 QTL (95th percentile threshold) and two QTL (99th percentile threshold). To account for false negatives, we scanned for QTL using a more permissive 75th percentile cutoff. Of the 63 single-environment QTL, only nine had QTL in other environments by this more permissive threshold. Thus, the majority of the 63 single-environment QTL most likely result from environmentally contingent genetic effects on the ionome.

QTL by Environment Interactions

That QTL detection was so strongly affected by environment suggested the effects of allelic variation on element concentration were heavily dependent on environmental variables. These results, however, did not specifically test for QTL by environment interactions (QEI). Comparison between environments with our data is additionally complicated because different subsamples of the IBM population were grown at these multiple locations and years. While there are many different approaches to identifying QEI described in the literature (summarized in El-Soda et al. [18]) we focused on two previously implemented methods. The first considered

location (but not year) by comparing the goodness of fit for linear models with and without an interactive covariate [19–21]. The second method takes advantage of the ability to grow the same RILs in multiple years. Trait values measured in the same IBM line for the same element at the same site but in different years were subtracted from each other and the difference between years was assigned as the trait value for that RIL genotype for QTL detection [22, 23].

Linear model estimation of QTL by location effects. The most common approach to analyze QEI is to fit a linear model with environment as both a cofactor and an interactive covariate and compare results to a model with environment as an additive covariate [24]. This method is most amenable when data are available for the same lines grown in every environment, which was not the case across all of our dataset. Data from the three locations with two replicate years each (FL, IN, NY) were analyzed to reduce the number of covariates and increase the power to detect variation from the environment. The data for both years in each location were combined (FL05 & FL06, IN09 & IN10, NY05, NY06 & NY12), designating covariates based on location.

Two linear QTL models were fit to the combined data using the FL, IN, and NY locations as covariates. These models reflect the dependence of phenotype on genotype, environment, and genotype-by-environment interactions.

$$y_i = \mu + \beta_g g_i + \beta_x x_i + \gamma g_i x_i + \varepsilon_i \qquad (1)$$

$$y_i = \mu + \beta_g g_i + \beta_x x_i + \varepsilon_i \qquad (2)$$

The first equation fit (1) is the full model considering the phenotype of individual i (y_i) as controlled by genotype (g_i), location (x_i), and genotype by location interaction ($g_i x_i$), while the reduced model (2) estimates phenotype without considering genotype by location interaction, using genotype and location as purely additive factors. B_g and B_x represent the additive effects of

genotype and environment, respectively, while γ represents the effect of the genotype by environment interaction. By using likelihood ratio tests on full and reduced models, we can test the hypothesis that genotype by environment interactions significantly improve the fit of the model to the data and estimate the effects of genetic by environment interactions.

The program R/qtl was used to fit QTL using both the full and reduced models for sample weight and 20 elements, with three locations encoded as covariates in the environment term. For each marker, LOD scores resulting from the reduced QTL model were subtracted from LOD scores determined by the full model, leaving a LOD score for each marker representing solely the significance of the genetic by location component. The significance threshold for the subtracted LOD scores was calculated by using 1000 permutations of the three step procedure (fitting the two models with randomized data and then subtracting LOD scores). Even with this underpowered dataset, 10 QTL by location interactions exceeded the threshold (α= 0.05, Table 2). Interactions between QTL and location are likely to be due to a combination of soil and weather differences across different locations. In the case of Ni, our initial single-element QTL mapping conducted separately on data from each environment identified differences in QTL presence or strength between FL, IN, and NY locations for a QTL located at the beginning of chromosome 9 (Fig 2). This QTL corresponds to a locus found to have a significant QTL by location effect (Table 2). Remarkably, all elemental QTL by location interactions detected by this approach affected trace element accumulation. These elements are both low in concentration in the grain, and often variable among soils [25]. Cd, an element for which we found significant QEI, has detrimental effects on both human and plant health [26] and is toxic in food at levels as low as .05 ppm. [27]. The locus with the strongest QEI for Cd does not follow location averages of Cd content in the grain (Table S3) and therefore is unlikely to be affected by crossing a

detection threshold driven by higher Cd in the soils at those locations. The lack of direct correlation between QTL significance and grain content also occurs for the loci with strong by-location effects for Mo and Ni. This demonstrates that reduced cadmium or enhanced micronutrient contents in grain require plant breeding selections that consider complex genetic by environment interactions rather than genotypes assessed in a single soil environment.

Table 2. QTL with Significant by-Location interactions.

Trait	Chr	Pos (cM)	LOD	Significance Threshold†
Mn	1	232.4	7.03	4.59
Mn	5	195.8	4.61	4.59
Fe	5	204.6	4.50	3.94
Ni	1	410.3	6.15	4.69
Ni	9	7.7	28.50	4.69
Cu	7	165.9	5.31	4.72
Zn	4	157.4	4.44	4.13
Rb	2	185.3	3.38	2.80
Mo	1	378.0	48.49	4.20
Cd	2	214.6	20.26	3.87

†α= 0.05

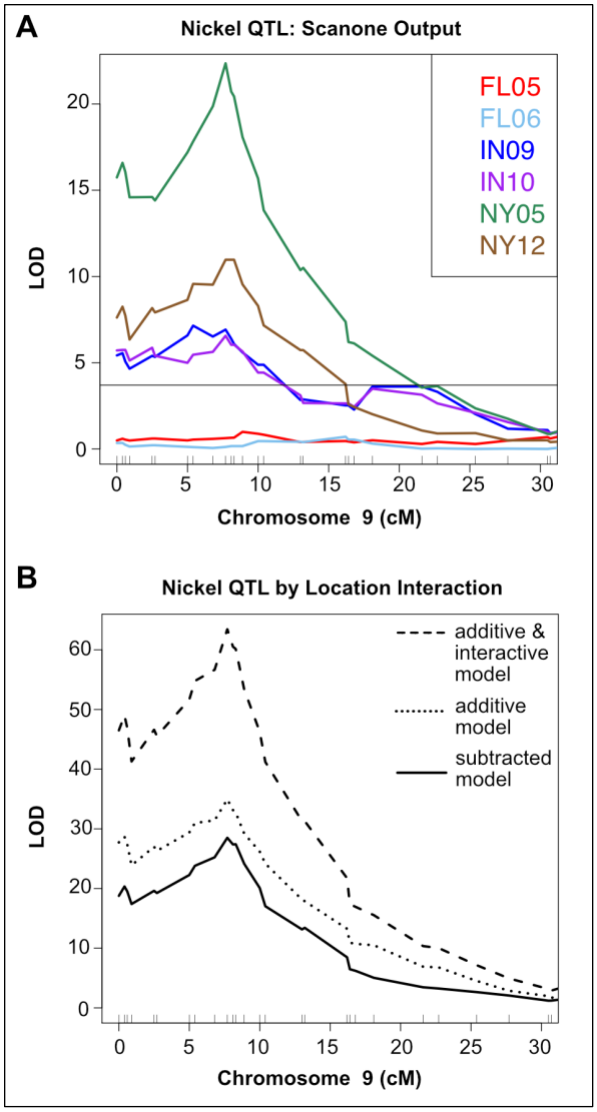


Fig 2. Significant QTL-by-Location Interactions Reflect Variation in Single Environment Mapping. (A) Nickel QTL on chromosome 9 exhibits variation between FL, IN, and NY growouts in single environment QTL mapping. Scanone QTL mapping output for Ni on a segment of Chr9 is plotted for FL05, FL06, IN09, IN10, NY05, and NY12. LOD score is plotted on the y-axis and cM position on the x-axis. Horizontal line corresponds to significance threshold (α= 0.05). (B) Scanone QTL mapping for combined Ni data from Florida (FL05 and FL06), Indiana (IN09 and IN10), and New York (NY05 and NY12) growouts. All lines within each location were included, with covariates designated based on location. QTL mapping output using model with location as an additive covariate is shown as dotted line. QTL mapping output from model with location as both an additive and interactive covariate is shown as dashed line. Subtracted LOD score profile from the two models (QTL by location interactive effect only) is shown as solid line. Horizontal line corresponds to significance threshold for QTL by location interaction effect, derived from 1000 iterations of the three step procedure using randomized data: scanone QTL mapping with the additive model, scanone QTL mapping with the additive and interactive model, and subtraction of the two models.

QTL for trait differences within location. The previous method identified genotypes with interactions with location but not with year. Year to year variation will also have effects due to differences in rainfall, temperature and management practices. To examine variation that occurs within a location over different years, we examined intra-location QEI in the three previously used locations with two year samples (FL05 & FL06, IN09 & IN10, NY05 & NY12). QTL were mapped using the stepwise algorithm on trait differences for sample weight and 20 elements between common lines among the two different years from a location. This approach identified loci affecting phenotypic differences between the same lines grown on the same farm but in different years. Six QTL were found for FL05-FL06 differences, one QTL for IN09-IN10 differences, and two QTL for NY05-NY12 differences (Table 3). These trait-difference QTL included loci identified in our single element/environment QTL experiment where a locus was present for one year but not the other or the QTL was found in both years with differing strength (Fig 3A, B, C). Six of the difference QTL were detected at loci where no QTL were detected when the years were mapped separately, revealing novel gene by environment interactions not obvious from the single year data. These significant effects of year-to-year environmental variation within the same location indicated that factors beyond location are both influencing the ionome and determining the consequences of genetic variation.

Table 3. Significant QTL for Trait Differences.

Location	Years Compared	Trait	Chr	Pos (cM)	LOD	Significance Threshold[†]
FL	FL05_FL06	Mg	8	294.4	5.23	3.74
FL	FL05_FL06	P	4	130.2	3.89	3.60
FL	FL05_FL06	P	4	297.8	6.03	3.60
FL	FL05_FL06	P	8	294.6	8.43	3.60
FL	FL05_FL06	Co	1	296.3	4.36	3.69
FL	FL05_FL06	Mo	1	378.6	6.10	3.70
IN	IN09_IN10	Fe	8	140.9	4.52	3.62

NY	NY05_NY12	K	5	154.2	4.25	3.61
NY	NY05_NY12	Sr	7	193.2	4.45	3.66

†α= 0.05

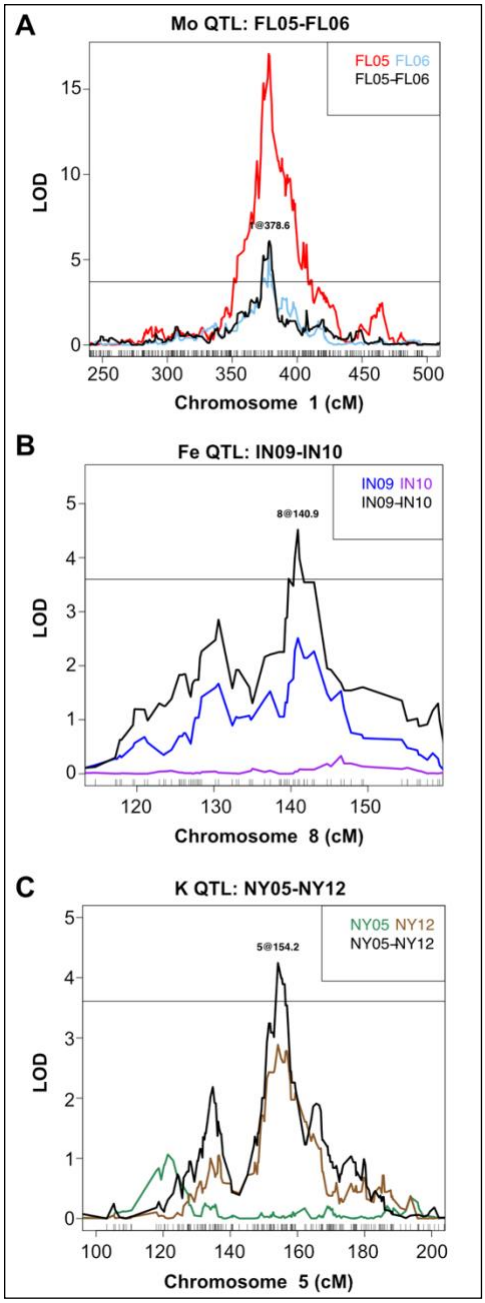


Fig 3. Comparison of QTL Mapped on Traits in Single Environments and Trait Differences Between Environments. Examples from stepwise QTL mapping on trait differences between two years at one location, calculated between IBM lines common to both years. Scanone QTL mapping output is also plotted for the same trait from each year separately. LOD score is shown on the y-axis and cM position on the x-axis. Horizontal lines correspond to

significance threshold (α= 0.05). (A) Molybdenum QTL on chromosome 1 mapped for Mo in FL05, Mo in FL06, and difference in Mo content between FL05 and FL06. (B) Iron QTL on chromosome 8 mapped for Fe in IN09, Fe in IN10, and difference in Fe content between IN09 and IN10. (C) Potassium QTL on chromosome 5 mapped for K in NY05, K in NY12, and difference in K content between NY05 and NY12.

DISCUSSION

The results described here demonstrate that the concentrations of elements in the kernels of maize are strongly affected by the interaction of genetics with growth environment. The majority of elements exhibited higher heritability within each environment and a dramatic drop in heritability across multiple environments. Combined with the presence of a large number of single-environment QTL, these data support the hypothesis that environment has a significant impact on genetic factors influencing the ionome. By changing the stringency of the statistical tests, we are able to discount the likelihood that that these single environment QTL are the result of a large number of false positives or false negatives. The structure of our data, with few lines measured across all locations and years, limited our ability to test for direct QTL by environment interactions. As a result, we have likely underestimated the extent of QEI. Future studies with uniform lines across environments will allow for inclusion of data from all environments and lines and increase power to detect additional genetic by environment interactions.

Nevertheless, we were able identify QEI over different locations and QEI at a single location over different years. We identified a strong nickel QTL on chromosome 9 that was found in Indiana and New York with single-environment QTL mapping, but not in Florida. This same locus also was found to be a significant location-interacting QTL when using a model that included Indiana, New York, and Florida as covariates. One possible cause for this, and other location specific QTL, might be differences in element availability between local soil

environments. Interestingly, the presence/absence of the QTL does not seem to correlate with the mean levels of the elements in the grains sampled from that location, suggesting that QEI are not being driven solely by altered availability of the elements in the soil. Local soil differences are less likely to be driving the QTL found for pairwise differences between two years at one location. Soil content should remain relatively similar from year to year at the same farm, suggesting that the loci identified by comparison between years and within location will encode components of elemental regulatory processes responsive to precipitation, temperature, or other weather changes. Experiments with more extensive weather and soil data, or carefully manipulated environmental contrasts, are needed to create models with additional covariates and precisely model environmental impacts.

Although the mapping intervals do not provide gene-level resolution, several QTL overlap with known elemental regulation genes, such as the QTL on chromosome 1 at 378 cM which coincides with ZEAMMB73_045160, an ortholog of the Arabidopsis molybdenum transporter, MOT1. We observe strong effects and replication of this QTL across nearly all environments, suggesting that this MOT1 ortholog plays a role in a variety of environments. Other large effect QTL found in several environments merit further investigation, as they may recapitulate important element-associated genes that have yet to be identified. Identification of the genes underlying these QTL and the gene/environmental variable pairs underlying the QEIs will improve our understanding of the factors controlling plant elemental uptake and productivity. Given the high levels of variability that the interaction between genotype and environmental factors can induce in these traits, conventional breeding approaches that look for common responses across many different environments for a single trait may fail to improve the

overall elemental content, necessitating rational approaches that include both genetic and environmental factors.

ACKNOWLEDGEMENTS

The authors would like to thank Justin Borevitz and Riyan Cheng, and Tom Juenger for advice on QTL by environment mapping and Karl Broman for invaluable assistance with R/qtl. The authors would especially like to thank our field collaborators Sherry Flint-Garcia, Peter Balint-Kurti, Torbert Rocheford, Jonathan Lynch, and Robert Snyder for their dedicated efforts to provide the seeds analyzed for this project.

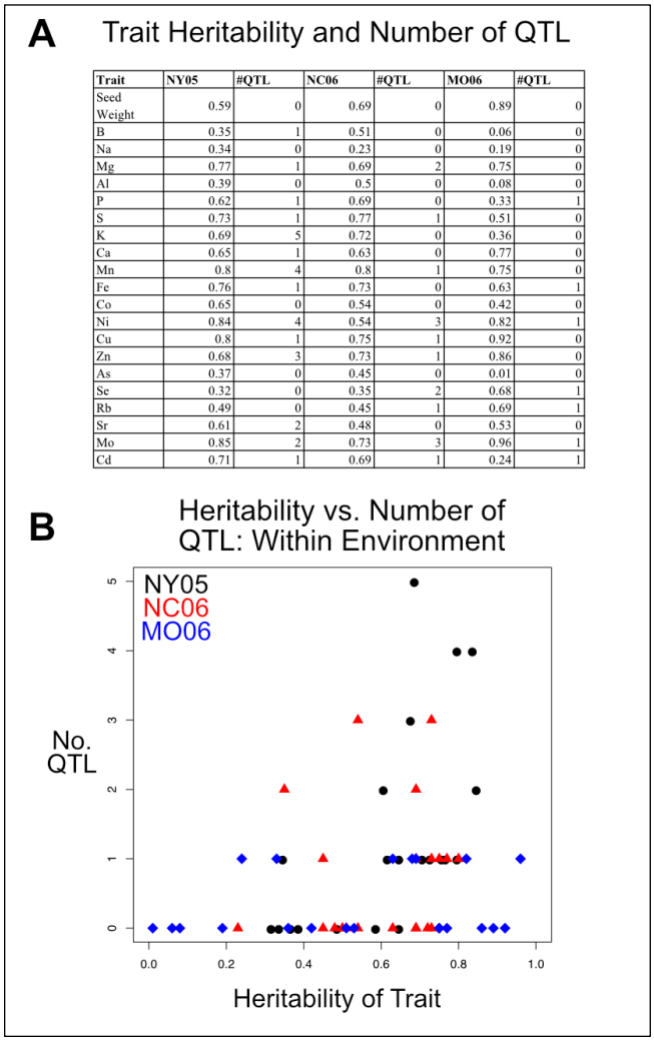
REFERENCES

1. McDowell SC, Akmakjian G, Sladek C, Mendoza-Cozatl D, Morrissey JB, Saini N et al. Elemental Concentrations in the Seed of Mutants and Natural Variants of *Arabidopsis thaliana* Grown under Varying Soil Conditions. PLoS ONE. 2013;8: e63014.
2. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet. 2013;126: 867-887.
3. Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE et al. Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana. Nat Biotechnol. 2003;21: 1215-1221.
4. Koornneef M, Alonso-Blanco C, Peeters AJM. Genetic approaches in plant physiology. New Phytologist. 1997;137: 1-8.
5. Ziegler G, Terauchi A, Becker A, Armstrong P, Hudson K, Baxter I. Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. The Plant Genome. 2013;6:
6. Davies L, Gather U. The Identification of Multiple Outliers. Journal of the American Statistical Association. 1993;88: 782-792.
7. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D et al. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant molecular biology. 2002;48: 453-461.
8. Balint-Kurti PJ, Zwonitzer JC, Wisser RJ, Carson ML, Oropeza-Rosas MA, Holland JB et al. Precise Mapping of Quantitative Trait Loci for Resistance to Southern Leaf Blight,

Caused by Cochliobolus heterostrophus Race O, and Flowering Time Using Advanced Intercross Maize Lines. Genetics. 2007;176: 645-657.
9. Dubois PG, Olsefski GT, Flint-Garcia S, Setter TL, Hoekenga OA, Brutnell TP. Physiological and Genetic Characterization of End-of-Day Far-Red Light Response in Maize Seedlings. Plant Physiol. 2010;154: 173-186.
10. Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD. Quantitative Trait Loci and Comparative Genomics of Cereal Cell Wall Composition. Plant Physiology. 2003;132: 263-271.
11. Lung'aho MG, Mwaniki AM, Szalma SJ, Hart JJ, Rutzke MA, Kochian LV et al. Genetic and physiological analysis of iron biofortification in maize kernels. PLoS One. 2011;6: e20429.
12. Ordas B, Malvar RA, Santiago R, Butron A. QTL mapping for Mediterranean corn borer resistance in European flint germplasm using recombinant inbred lines. BMC Genomics. 2010;11: 1-10.
13. Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Höhne M et al. Fine Quantitative Trait Loci Mapping of Carbon and Nitrogen Metabolism Enzyme Activities and Seedling Biomass in the Intermated Maize IBM Mapping Population. Plant Physiology. 2010.
14. Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, Connolly EL et al. Integration of experiments across diverse environments identifies the genetic determinants of variation in Sorghum bicolor seed element composition. Plant physiology. 2016; pp-01971.
15. Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64: 641-656.
16. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138: 963-971.
17. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J et al. Variation in Molybdenum Content Across Broadly Distributed Populations of *Arabidopsis thaliana* Is Controlled by a Mitochondrial Molybdenum Transporter (*MOT1*). PLoS Genet. 2008;4: e1000004.
18. El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. Genotype x environment interaction QTL mapping in plants: lessons from Arabidopsis. Trends in Plant Science. 2014;19: 390-398.
19. Bhatia A, Yadav A, Zhu C, Gagneur J, Radhakrishnan A, Steinmetz LM et al. Yeast Growth Plasticity Is Regulated by Environment-Specific Multi-QTL Interactions. G3. 2014;4: 769-777.
20. Leinonen PH, Remington DL, Leppälä J, Savolainen O. Genetic basis of local adaptation and flowering time variation in Arabidopsis lyrata. Molecular ecology. 2013;22: 709-723.
21. Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA, Thorgaard GH. Quantitative Trait Loci x Maternal Cytoplasmic Environment Interaction for Development Rate in Oncorhynchus mykiss. Genetics. 2007;175: 335-347.
22. Tétard-Jones C, Kertesz MA, Preziosi RF. Quantitative trait loci mapping of phenotypic plasticity and genotype-environment interactions in plant and insect performance. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2011;366: 1368-1379.
23. Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TFC. Genotype-Environment Interactions at Quantitative Trait Loci Affecting Inflorescence Development in Arabidopsis

thaliana. Genetics. 2003;165: 353-365.
24. Des Marais DL, Hernandez KM, Juenger TE. Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annual Review of Ecology, Evolution, and Systematics. 2013;44: 5-29.
25. White JG, Zasoski RJ. Mapping soil micronutrients. Field Crops Research. 1999;60: 11-26.
26. Godt J, Scheidig F, Grosse-Siestrup C, Esche V, Brandenburg P, Reich A et al. The toxicity of cadmium and resulting hazards for human health. J Occup Med Toxicol. 2006;1: 22.
27. United States Department of Agriculture GAIN. China's Maximum Levels of Contaminants in Food. GAIN report CH14058. 2014

SUPPORTING INFORMATION



Trait	NY05	#QTL	NC06	#QTL	MO06	#QTL
Seed Weight	0.59	0	0.69	0	0.89	0
B	0.35	1	0.51	0	0.06	0
Na	0.34	0	0.23	0	0.19	0
Mg	0.77	1	0.69	2	0.75	0
Al	0.39	0	0.5	0	0.08	0
P	0.62	1	0.69	0	0.33	1
S	0.73	1	0.77	1	0.51	0
K	0.69	5	0.72	0	0.36	0
Ca	0.65	1	0.63	0	0.77	0
Mn	0.8	4	0.8	1	0.75	0
Fe	0.76	1	0.73	0	0.63	1
Co	0.65	0	0.54	0	0.42	0
Ni	0.84	4	0.54	3	0.82	1
Cu	0.8	1	0.75	1	0.92	0
Zn	0.68	3	0.73	1	0.86	0
As	0.37	0	0.45	0	0.01	0
Se	0.32	0	0.35	2	0.68	1
Rb	0.49	0	0.45	1	0.69	1
Sr	0.61	2	0.48	0	0.53	0
Mo	0.85	2	0.73	3	0.96	1
Cd	0.71	1	0.69	1	0.24	1

Fig S1. Heritability vs. Number of QTL. (A) Comparison between broad-sense heritability (H^2) of seed weight and elemental traits. In environments with >1 replicate (NY05, NC06, MO06), H^2 was calculated for each trait. Table indicates H^2 and number of QTL detected for

each trait in the designated environment. (B) Plot of trait heritability vs. the number of QTL identified in the respective environment. Environments are indicated by color.

Table S1. Growout Information.

For PCA: post-OR is OR after removing poorly measured elements

Location	Year	Planting Date	No. Lines	No. Line Reps.*	Genotyped Lines†	No. Lines Post-OR	Genotyped Lines Post-OR
Florida	2005	9/14/2005	220	1(118), 2(2)	176	180	147
Florida	2006	8/25/2006	118	1(71), 2(47)	95	114	94
Indiana	2009	5/9/2009	193	1	156	169	134
Indiana	2010	5/10/2010	168	1	139	155	129
North Carolina	2006	5/6/2006	197	1(19), 2(121), 3(53), 4(4)	160	187	151
New York	2005	5/9/2006	256	1(7), 2(50), 3(199)	209	249	204
New York	2006	5/9/2006	82	1(60), 2(22)	67	56	46
New York	2012	5/24/2012	168	1	137	128	104
Missouri	2006	5/17/2006	97	1(29), 2(50), 3(18)	81	58	50
South Africa	2010	11/2009	88	1	72	82	68

*No. lines with rep. in parentheses

†239 total genotyped lines

Lines with any elemental outliers were removed prior to PCA.

Table S2. Percent Variance (R^2) of Mo, Cd, and Ni QTL.

	Mo 1@378	Cd 2@215	Ni 9@7
FL05	33.99	43.36	NA
FL06	27.13	27.08	NA
IN09	26.85	38.65	21.85
IN10	33.35	44.77	19.67
NC06	31.95	48.88	32.01
NY05	69.85	52.17	47.61
NY06	45.17	21.61	NA
NY12	57.19	60.44	35.10
MO06	58.21	NA	NA
SA10	NA	NA	NA

Percent variance for 3 QTL in locations where QTL is significant. QTL chromosome and position is indicated under element name.

Table S3. Location LOD Scores Compared to Seed Element Content.

	FL	IN	NY
Cd_2@214_LOD	16.96	23.81	37.13
Cd_2@214_normalizedLOD	0.08	0.15	0.17
Avg_Cd	0.42	0.44	0.21
Mo_1@378_LOD	11.31	17.64	51.72
Mo_1@378_normalizedLOD	0.06	0.11	0.24
Avg_Mo	3.22	4.85	1.99
Ni_9@7.7_LOD	0.47	8.12	23.25
Ni_9@7.7_normalizedLOD	0.00	0.05	0.11
Avg_Ni	1.01	2.31	0.95

Comparison for top three significant QTL-by-location interaction loci (Cd, Mo, Ni)

CHAPTER 3:
MULTIVARIATE ANALYSIS REVEALS ENVIRONMENTAL AND GENETIC DETERMINANTS OF ELEMENT COVARIATION IN THE MAIZE GRAIN IONOME

Alexandra Asaro, Brian P. Dilkes, Ivan Baxter

This chapter is currently under peer review for publication.
AA conducted all data analysis. AA interpreted results, created figures, and wrote the manuscript with assistance from BD and IB.

ABSTRACT

The integrated responses of biological systems to genetic and environmental variation results in substantial covariance in multiple phenotypes. The resultant pleiotropy, environmental effects, and genotype-by-environment interactions (GxE) are foundational to our understanding of biology and genetics. Yet, the treatment of correlated characters, and the identification of the genes encoding functions that generate this covariance, has lagged. As a test case for analyzing the genetic basis underlying multiple correlated traits, we analyzed maize kernel ionomes from Intermated B73 x Mo17 (IBM) recombinant inbred populations grown in 10 environments. Plants obtain elements from the soil through genetic and biochemical pathways responsive to physiological state and environment. Most perturbations affect multiple elements which leads the *ionome*, the full complement of mineral nutrients in an organism, to vary as an integrated network rather than a set of distinct single elements. We compared quantitative trait loci (QTL) determining single-element variation to QTL that predict variation in principal components (PCs) of multiple-element covariance. Single-element and multivariate approaches detected partially overlapping sets of loci. QTL influencing trait covariation were detected at loci that were not found by mapping single-element traits. Moreover, this approach permitted testing environmental components of trait covariance, and identified multi-element traits that were determined by both genetic and environmental factors as well as genotype-by-environment interactions. Growth environment had a profound effect on the elemental profiles and multi-element phenotypes were significantly correlated with specific environmental variables.

INTRODUCTION

Elements are distinct chemical species, and studies of element accumulation frequently investigate each element separately. There is overwhelming evidence, however, that element accumulations covary due to physical, physiological, genetic, and environmental factors. In a dramatic example in *Arabidopsis thaliana*, a suite of elements responds to Fe deficiency in such a concerted manner that they can be used to predict the deficiency or sufficiency of Fe for the plant more accurately than the measured level of Fe in plant tissues [1]. The basis of this covariation can be as simple as co-transport of multiple elements. IRT1 is a metal transporter capable of transporting Fe, Zn, and Mn. IRT1 is upregulated in low Fe conditions resulting in an environmentally-dependent link between Fe and other ions [2]. Other pairs of co-regulated elements, such as Ca and Mg which share homeostatic pathways in *Brassica oleracea* [3], should be affected predictably by genetic variation. When *A. thaliana* recombinant inbred line populations were grown in multiple environments, genetic correlations among Li-Na, Mg-Ca, and Cu-Zn were observed across all environments while Ca-Fe and Mg-Fe were only correlated in a subset of environments [4]. Shared genetic regulation of ion transport without substantial environmental responsiveness should result in the former pattern, along with significantly less capacity for homeostasis across environmental concentrations and availabilities of elements. Environmentally-responsive molecular mechanisms, reminiscent of *IRT1* upregulation, could result in environmentally-variable patterns of correlations. Baxter et al. previously tested element seed concentrations for correlations in the maize Intermated B73 x Mo17 (IBM) recombinant inbred population, finding several correlated element pairs, the strongest of which was between Fe and Zn [5]. Yet, few QTL impacting more than one element were found, possibly due to QTL with small effects on multiple elements failing to exceed the threshold of observation when

mapping on single element traits with limited numbers of lines. Thus, while understanding the factors driving individual element accumulation is important, we must consider the ionome as a network of co-regulated and interacting traits [6]. We propose that formally considering this coordination between elements can provide deeper insight than focusing on each element in isolation and that this will be a general feature of massively parallel phenotyping data and homeostatic systems.

Multivariate analysis techniques, such as principal components analysis (PCA), can reduce data dimension and summarize covariance of multiple traits as vectors of values by minimizing the variances of input factors to new components. When multiple phenotypes covary, as occurs for the elements in the ionome, this approach may complement single element approaches by describing trait relationships. In studies on crops such as maize, PCA has been used as a strategy to consolidate variables that may be redundant or reflective of a common state [7–11]. PCA has proved useful in previous QTL mapping efforts, facilitating detection of new PC QTL not found using univariate traits in analyses of root system architecture in rice [12] and kernel attributes, leaf development, ear architecture, and enzyme activities in maize [13–15]. In the current study, we expect that elemental variables are functionally related and therefore need new traits to describe elemental covariation. Since we do not know the exact nature of these relationships, and the ionome varies depending on environment, PCA is useful in that it does not require *a priori* definition of relationships between variables. If the PCA approach leads to novel loci and insights into how the ionome is functioning, it will be a valuable addition to the study of mineral nutrient regulation.

Here we show that developing multivariate traits reveals environmental and genetic effects that are not detected using single elements as traits. We performed PCA on element

profiles from the maize IBM population [16] grown in 10 different environments. Different relationships between elements were identified that depended on environment. QTL mapping using multi-element PCs as traits was carried out within each environment separately. Comparing these multivariate QTL mapping results to previous single-element QTL analyses of the same data [17] demonstrates that a multivariate approach uncovers unique loci affecting multi-element covariance. Additionally, experiment-wide PCA performed on combined data from all environments produced components capable of separating lines by environment based on their whole-ionome profile. These experiment-wide factors, while representative of environmental variation, also exhibited a genetic component, as loci affecting these traits were detected through QTL mapping. This shared involvement in element covariation is the expectation of genetic and environmental variation resulting in adjustments to the physiological mechanisms underlying adaptation.

MATERIALS & METHODS

Field Growth and Data Collection

Field growth and elemental profile analysis. Lines belonging to the Intermated B73 x Mo17 recombinant inbred (IBM) population [16] were grown in 10 different environments: Homestead, Florida in 2005 (220 lines) and 2006 (118 lines), West Lafayette, Indiana in 2009 (193 lines) and 2010 (168 lines), Clayton, North Carolina in 2006 (197 lines), Poplar Ridge, New York in 2005 (256 lines), 2006 (82 lines), and 2012 (168 lines), Columbia, Missouri in 2006 (97 lines), and Ukilima, South Africa in 2010 (87 lines). Elemental analysis was carried out in a standardized inductively coupled plasma mass spectrometry (ICP-MS) pipeline previously

described in detail [17]. Analytical outlier removal and weight normalization was performed following data collection as described in our previous analysis of these data.

Computational Analysis

Element correlation analysis. Within environments, 190 Pearson correlation coefficients were calculated, one for each pair of the 20 measured elements. To control for multiple tests, we applied a Bonferroni correction at an alpha level of 0.05. Given 190 possible combinations, correlations with a p-value below 0.05/190 = 0.00026 were regarded as significant.

Principal components analysis of ionome variation within environments. Elements prone to analytical error (B, Na, Al, As) were removed before to PC analysis, leaving 16 elements: Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, Se, Rb, Sr, Mo, and Cd. B, Na, Al, and As have a fairly low signal to noise ratio; because all elements are scaled together in a PCA, including these elements would increase the amount of noise variation going into the PCA. In an attempt to summarize the effects of genotype on covariance of ionomic components, a PCA was done using elemental data for each of the 10 environments separately. The *prcomp* function in R with scale = TRUE was used for PCA on elemental data to perform PCA on the line average element values in an environment. This function performs singular value decomposition on a scaled and centered version of the input data matrix, computing variances with the divisor *N*-1. 16 PCs were returned from each environment. The IBM population is a large and diverse population and we observe extensive variation across the elements, so even a small proportion of variation could explain a substantial amount of actual variation. We used a PCA screeplot to guide our choice of a 2% cutoff (Fig S1). After removal of PCs accounting for less than 2% of

the variance, the 10 sets of PCs were used as traits in QTL analysis. Variance proportions and trait loadings for all PCs calculated across 10 environments are provided in Table S3.

QTL Mapping: principal components. QTL mapping was done using stepwise forward-backward regression in R/qtl [34] as described previously for element phenotypes [17]. The mapping procedure was done for each environment separately, with PC line means for RILs in the given environment as phenotypes and RIL genotypes as input. The *stepwiseqtl* function was used to produce an additive QTL model for each PC, with the max number of QTL allowed for each trait set at 10. The 95th percentile LOD score from 1000 *scanone* permutations was used as the penalty for addition of QTL. 1000 permutations were done for every trait-environment combination. The QTL model was optimized using *refineqtl* for maximum likelihood estimation of QTL positions. The locations of the PC QTL detected in this study were compared to the single element QTL from our previous study. Loci were considered distinct if they were at least 25 cM away from any single element QTL detected in the environment in which the PC QTL was detected. This serves as a conservative control in order to minimize the mistaken assessment of novelty for QTL with small changes in peak position.

QTL by environment analysis: PCA across environments. The 16 most precisely measured elements were used for an additional principal components analysis. Again, the *prcomp* function in R with scale = TRUE was used for PCA on elemental data, however, all 16 element measurement values in all lines in all of the 10 environments were combined into one PCA. These PCs are referred to as across-environment PCs (aPCs). Element loadings were recorded and plotted along with lines colored by environment for aPCs 1 and 2 (Fig S4). The first 7 aPCs explained 93% of the total covariation of these traits. A linear model was used to test the relationship of environmental parameters on these aPCs. All seven aPCs were also used for

stepwise QTL mapping by the same method described above, with 1000 permutations for every trait-environment combination used to set 95th percentile significance thresholds.

QTL by environment analysis: Projection-PCA across environments. The sets of lines grown in each our ten environments were drawn from the same population [16] but different subsets were grown and harvested in different environments. To achieve common multivariate summaries for all lines and growouts, we performed an alternative PCA using a smaller set of common lines. We then projected the loadings from this PCA onto the full dataset, as follows. First, a PCA was conducted on 16 lines common to six of the 10 environments (FL05, FL06, IN09, IN10, NY05, NY12). The loadings for each PC from this PCA were then used to calculate values from full set of lines across 10 environments to generate PCA projections (PJs). These derived values based on a common-line PCA were compared to previously described aPC values from the PCA done on all lines at once. Correlations between PJs and aPCs were computed to compare the outcomes of the two methods.

Weather and soil data collection and analysis. Weather data for FL05, FL06, IN09, IN10, NC06, NY05, NY06, and NY12 was downloaded from Climate Data Online (CDO), an archive provided by the National Climatic Data Center (NCDC) through the National Oceanic and Atmospheric Administration (http://www.ncdc.noaa.gov/cdo-web/). Data were not available for the South Africa growout. Daily summary data for each day of the growing season were tabulated from the weather station nearest to the field location. Weather stations used to obtain data for each location are indicated in Table S4. Minimum temperature (in degrees Celsius) and maximum temperature (in degrees Celsius) were available in each location. With these variables, average minimum temperature, and maximum temperature were calculated across the 120-day growing season as well as for 30 day quarters. Growing degree days (GDD) were calculated for

the entire season and quarterly using the formula GDD = ((Tmax + Tmin)/2) – 10. No max or min thresholds were used in the GDD calculation.

Data describing soils from each location were obtained from the Web Soil Survey provided by the USDA Natural Resources Conservation Service (http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm). A representative area of interest was selected at the site of plant growth using longitude and latitude coordinates. When an area contained more than one soil type, a weighted average of measurements from all soil types was used. The data we downloaded from the Web Soil Survey were: pH, electrical conductivity (EC) (decisiemens per meter at 25 degrees C), available water capacity (AWC) (centimeters of water per centimeter of soil), available water supply (AWS) (centimeters), and calcium carbonate (CaCO3) content (percent of carbonates, by weight). Layer options were set to compute a weighted average of all soil layers.

The relationships between the seven experiment wide aPCs and the weather and soil variables were estimated by calculating Pearson correlation coefficients for the pairwise relationships. Correlations were also calculated between average element values and soil and weather variables in each environment.

RESULTS

Summary of Data Collection and Previous Analysis of Single Element Traits

We previously acquired data on 20 elements measured in the seeds from *Zea mays* L. Intermated B73 x Mo17 recombinant inbred line (IBM) populations [16] grown in 10 different location/year settings [17]. This work is briefly summarized here as it serves as the basis of our comparison. The kernels came from RILs of the IBM population cultivated across six locations

and five years. Quantification of the accumulation of 20 elements in kernels was done using inductively coupled plasma mass spectrometry (ICP-MS). Weight-adjusted element measurements were used for a QTL analysis to detect loci contributing to variation in seed element contents [17]. The current study is motivated by previous demonstrations of elemental correlations and mutant phenotype analyses which indicate extensive relationships between elements [1, 4]. To explore this formally, we further analyzed these data from a multiple-element perspective.

Element to Element Correlations

Several elements were highly correlated across the dataset, exhibiting pairwise relationships among lines in a given environment that passed a conservative Bonferroni correction for multiple tests. Many of these correlations reflected results previously obtained by Baxter et al., such as the strong correlation between Fe and Zn [5]. We detected 209 pairs of elements that were genetically correlated out of 1,900 possible correlations across environments (190 pairs per environment). We expect robust genetic influence to produce repeated observation of trait correlations in multiple environments. Of the six locations included in this experiment, we obtained data from three locations (FL, IN, and NY) from plant material grown in two different years. Seven element-pairs were correlated in five or more of these six environments: Mn and Mg, Mg and S, Mg and P, S and P, P and K, Ca and Sr, and Fe and Zn (Fig 1). Other element-pair correlations were driven by the genetic variation between IBM RIL in fewer environments. For example, Mn and P were correlated in FL05, NY05, and NY12 (r_p = 0.50, 0.48, 0.51) but were not significantly correlated in FL06, IN09, or IN10 (r_p = 0.31, 0.20, 0.18). Thus, while some correlations exist in multiple years and multiple locations, element correlations were affected by both location and year.

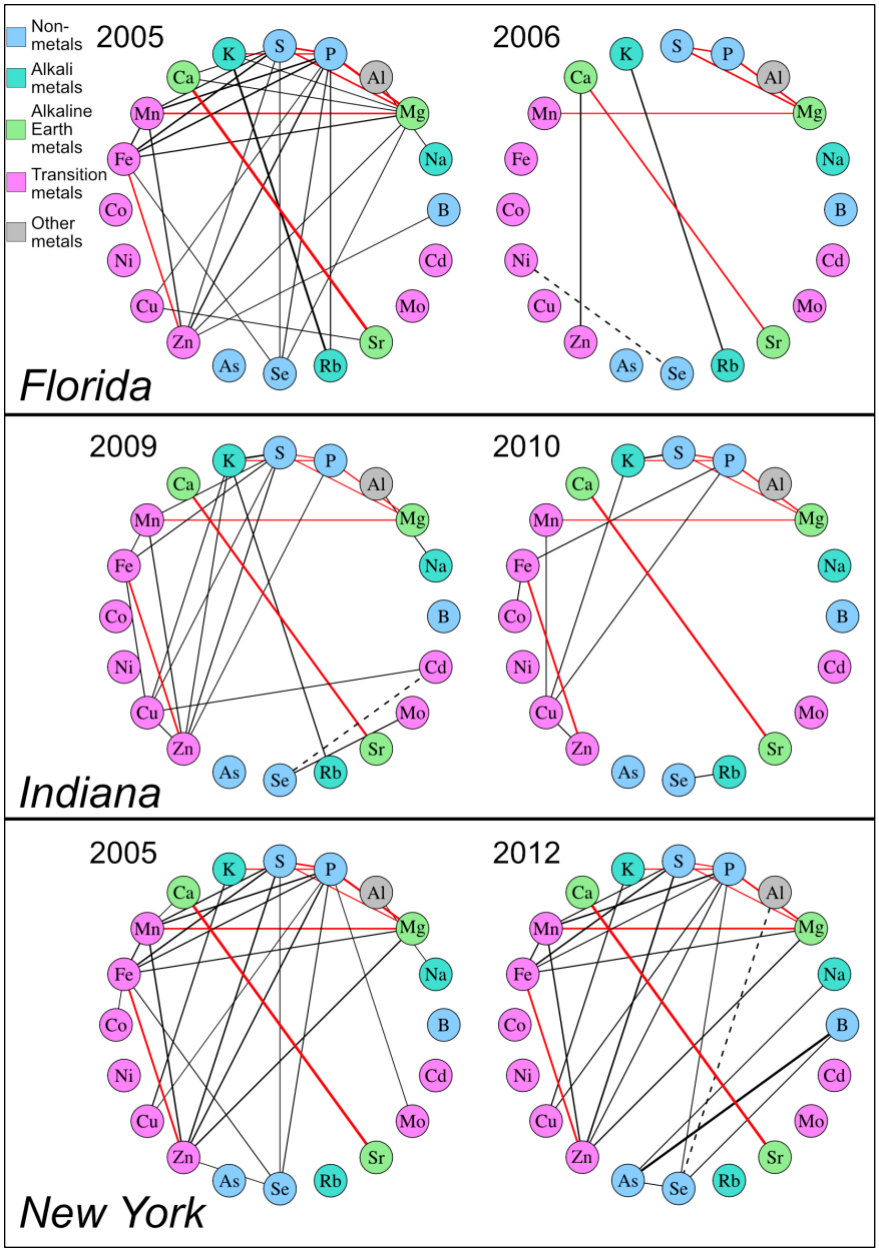


Fig 1. Element Correlations Diagrams for Locations with Repeated Measurements. Pairwise correlations of 20 kernel elements in varying environments, shown for the experiments within locations having data from multiple years (FL, IN, and NY). Correlations were calculated as the Pearson correlation coefficient (r_p) between concentration values for each element pair. Significance was evaluated using a Bonferroni correction for multiple tests within each environment and set at a corrected p value of 0.05. Lines between elements represent significant pairwise correlations, weighted by strength of correlation. Positive and negative correlations are represented as solid and dashed lines, respectively. Red lines indicate correlations present in at least 5 of the 6 environments shown.

In our previous single-element QTL analysis of these data, loci comprising QTL for two or more different elements were detected (Table 1). This mutual genetic regulation of multiple elements was readily apparent in the trait correlations calculated within environments, as five of the nine shared-element QTL exhibited corresponding element pair correlations within the given environment. For example, phosphorous, which was in three of the seven most reproducible element-pair correlations, exhibited the highest incidence of shared QTL with other elements. These included common QTL between P accumulation and all three of the reproducibly P-correlated elements: S and the cations K and Mg. In addition, P was affected by the only QTL shared between more than two elements, which affected P, S, Fe, Mn, and Zn accumulation in NY05 (Fig 2). Consistent with the possibility of variation in transport processes affecting element accumulation correlations, overlapping QTL were frequently found between elements with similar structure, charge, and/or type, such as Ca and Sr or Fe and Zn. These element correlations and post-hoc comparisons of shared QTL localizations suggest a genetic basis for covariance of the ionome in the RIL population.

Table 1. Loci Affecting Variation for Multiple Elements in the Same Environment.

Environment	Chr	Pos (cM) †	El 1	El 2	El 3	El 4	El 5
NY05	1	400	Mn	Ni	---	---	---
NY05	3	323	Sr	Ca	---	---	---
NY05	5	201	Mn	Zn	P	S	Fe
NY06	1	532	Mn	Mg	---	---	---
IN09	4	306	Fe	K	---	---	---
IN10	2	213	Mo	Cd	---	---	---
NY12	5	203	Zn	Fe	---	---	---
FL05	1	230	B	Mn	---	---	---
FL05	4	159	Fe	Zn	---	---	---

†Average position

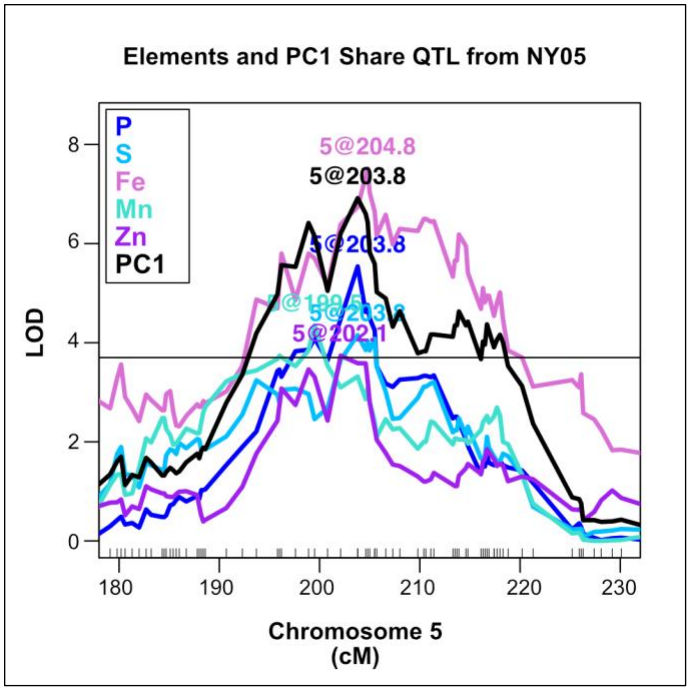


Fig 2. Multiple Element QTL. Stepwise QTL mapping output from the NY05 population for P, S, Fe, Mn, Zn, and PC1. Position in cM on chromosome 5 is plotted on the x-axis and LOD score is shown on the y-axis. 95th percentile of highest LOD score from 1000 random permutations is indicated as horizontal line.

Principle Components Analysis of Covariance for Elements in the Ionome

To better describe multi-element correlations and thereby detect loci controlling accumulation of two or more elements, we derived summary values representing the covariation of several elements. We implemented an undirected multivariate technique, principal components analysis, for this purpose. PCA reduced co-varying elements into principal components (PCs), orthogonal variables that account for variation in the original dataset, each having an associated set of rotations (also known as loadings) from the input variables. After removing elements prone to analytical artifacts, PCA was conducted using the remaining 16 elements from each of the 10 environments separately. This produced 16 principal components in each environment (Fig S1) of which we retained for further analysis only PCs representing

more than 2% of the total variation. This resulted in as few as 11 and as many as 15 PCs depending on environment.

Remarkably, there is substantial overlap in the loadings of many elements in the first and second PCs across some environments, suggesting a reproducible effect of genetic variation on the covariance of elemental accumulation in these environments (Fig 3). Additionally, the loadings of elements are consistent with the pair-wise relationships observed in the element-by-element correlations. For example, the chemical analogs Ca and Sr frequently load PCs in a similar direction. The PC loadings derive from inputs of several elements to a single PC variable. All retained PCs in all 10 environments have a loading contribution of at least 0.25 in magnitude from two or more elements. While some patterns existed across environments, many PC loadings differed in both magnitude and direction according to environment. This suggests instability of element-pair correlations across the environments. We used correlation tests of element loadings to detect PCs stemming from shared biological processes in each environment. This identified PCs from each environment that were constructed from similar relationships. Because loading direction is arbitrary, both strong positive and strong negative correlations were examined. 52 pairs of PCs exhibited loadings correlations with a Pearson correlation coefficient greater than 0.75 or less than -0.75 (Fig S2). Thus, the PC analyses of data across from different locations described similar patterns of elemental covariation, while not necessarily recovered in the same rank order.

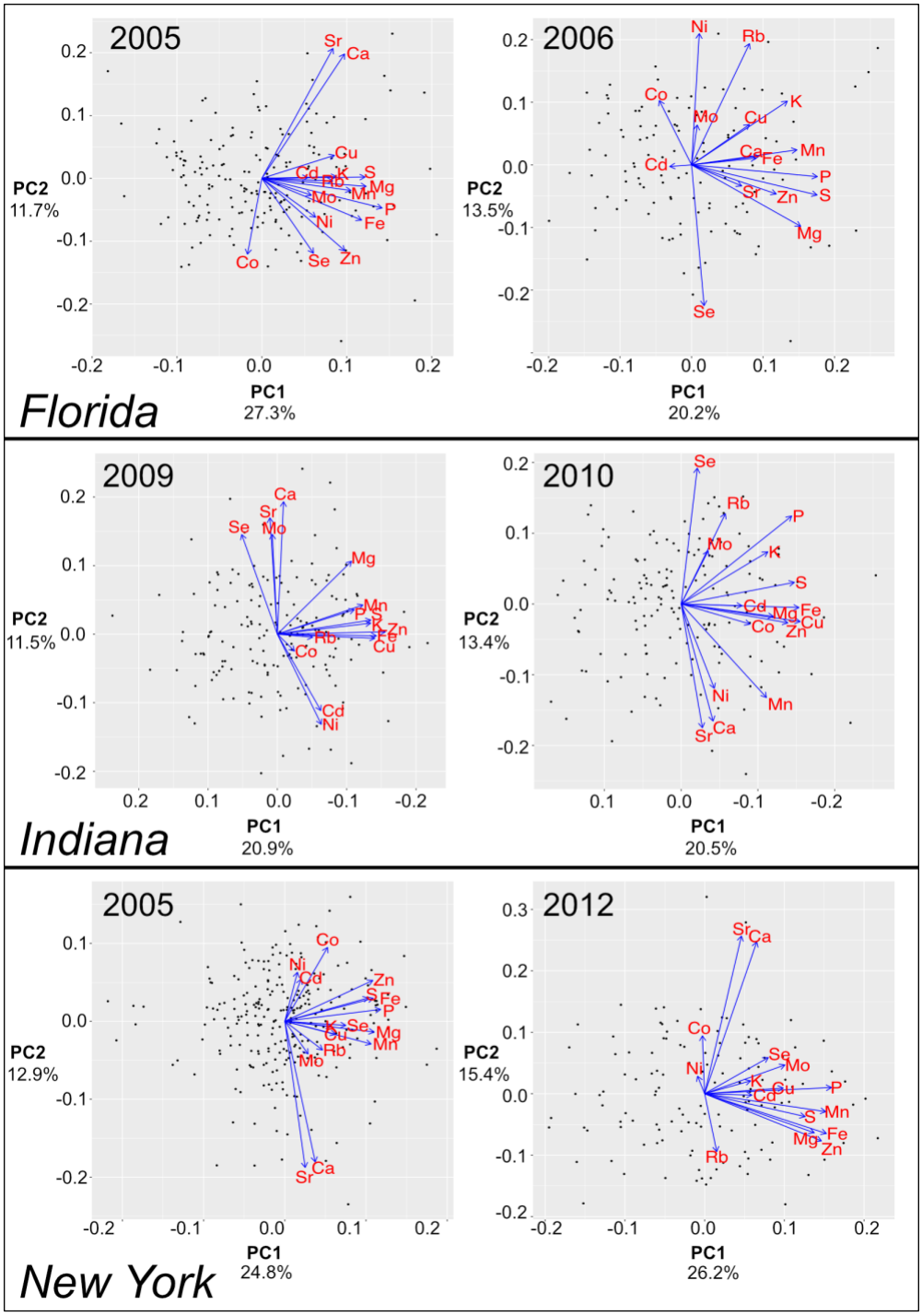


Fig 3. PCA Plots in Multiple Environments. PCA plots showing PC1 and PC2 loadings in different years in three locations (FL, IN, and NY). PC1 and PC2 values for each line are plotted as points and PC1 and PC2 loadings of each element are indicated by blue arrows. The data for different years for each of three locations, FL, IN, and NY are plotted. The percent of total variation explained by each PC is labeled on the axes. PC negative and positive values are arbitrary, so the Indiana x-axes are switched in direction to aid visual comparisons.

QTL Mapping of Ionomic Covariance Components

The PCs from each environment were used as traits for QTL detection. Stepwise QTL mapping using these derived traits yielded 93 QTL that exceeded an statistical threshold of α=0.05 estimated by 1000 permutations performed for each trait-environment combination (Fig 4C). QTL were found for PC traits explaining both large and smaller proportions of variation (Table S1). 56 of these QTL affecting multiple-element covariance components overlapped with previously detected single-element QTL in the same environment [17] (Fig 4A). In some cases, two or more PC traits within an environment resolved to one single-element QTL. This was observed particularly for elements with strong effect QTL, such as Mo, Cd, and Ni. For example, in IN10, PC2 and PC10 both have QTL that co-localize with the Cd QTL on chromosome 2. Likewise, in NY05, PC3, PC5, PC6, and PC9 all detect QTL coinciding with the large-effect Ni QTL on chromosome 9. Each of these PCs are comprised of varying loadings of Ni, along with other elements. This demonstrates that, although the relationship among elements described by each PC is distinct, a locus affecting a single-element can be detected due to loading of that element into more than one PC. This repeated detection of the same locations contributes to the higher number and proportion of detected PC QTL that were shared with element QTL (56/93) than element QTL that were shared with PC QTL (18/79), although the same genomic locations underlie this overlap.

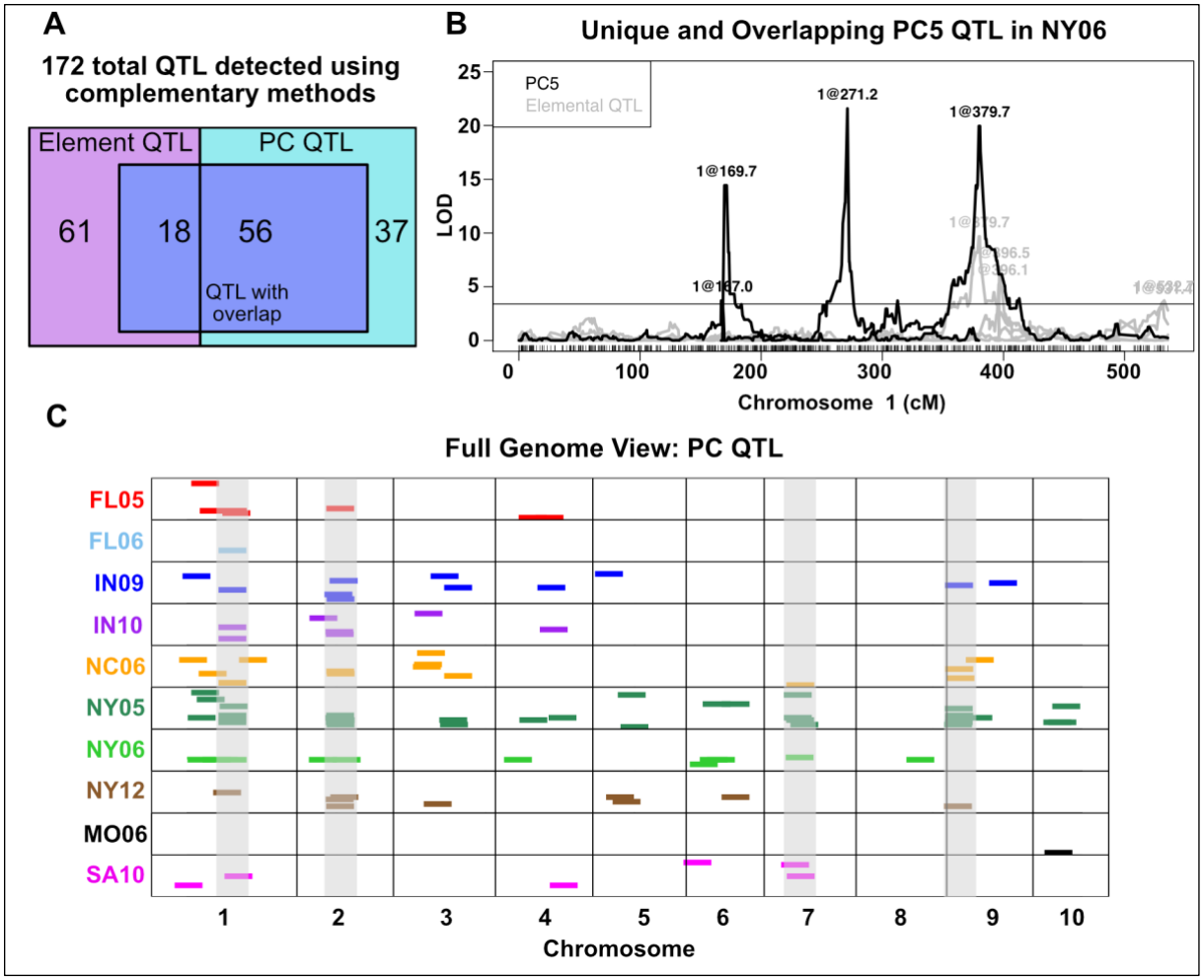


Fig 4. Principal Component QTL from 10 Environments. PCs were derived from elemental data separately in each of 10 environments and used as traits for QTL mapping. (A) 172 total element and PC QTL were mapped. The two boxes represent the 79 and 93 elemental and PC QTL, respectively. 18 element QTL overlap with PC QTL from the same environment. 56 PC QTL overlap with element QTL from the same environment. Sets of non-unique QTL are shown in the center box. QTL unique to elements, 61, and to PCs, 37, are shown outside of the shared box. (B) QTL mapping output for PC5 from the NY06 population. Position on chromosome 1 is shown on the x-axis, LOD score is on the y-axis. All significant NY06 element QTL on chromosome 1 are shown in grey ($\alpha = 0.05$). Two PC5 QTL, at 169.7 and 271.2 cM, are unique to PC5 and do not overlap with any elemental QTL. A PC5 QTL at 379.7 cM is shared with a molybdenum QTL. (C) Significant PC QTL ($\alpha = 0.05$) for PCs in 10 environments. QTL location is shown across the 10 chromosomes on the x-axis. Environment in which QTL was found is designated by color. QTL are represented as dashes of uniform size for visibility. Four regions highlighted in grey represent the four loci found for multiple PC traits in multiple environments (> 2).

QTL mapping on single elements may not have the power to detect loci with small coordinate effects on several elements. PC traits can reveal new QTL and enhance detection of common genetic factors modulating elements. 37 PC QTL were detected at loci not seen using single element traits. For instance, two PC5 QTL from the NY06 growout were located on chromosome 1 at positions distinct from any single-element QTL (Fig 4B). So as to not inflate PC-specific QTL, they are defined here as QTL greater than 25 cM away from any elemental QTL in the same environment. Top elemental loadings of PCs and overlap with elemental QTL is summarized in Table S2.

PC QTL analysis captured previously observed single-element QTL shared between elements within a particular environment. Of the nine loci affecting variation for multiple elements in the same environment (Table 1), four loci were detected for a PC trait in that environment (Table 2). For example, in NY05, a QTL was identified for PC1 that overlaps QTL that were detected in the single element analyses of P, S, Fe, Mn, and Zn on chromosome 5 (Fig 2). The log of odds score for this NY05 PC1 QTL was as strong as the association between the locus and Fe accumulation and more significant than the P, S, Mn, and Zn elemental QTL. Thus, the QTL for a multi-element PC was as strong as the best single-element approach for this previously detected multi-element locus. This is the prediction for traits that will affect variation in multiple elements, such as root structure or homeostatic processes. For these traits, the PC approach may be preferable to single elements, particularly in cases where single element changes are of small effect or below detection limits while concerted changes to multiple elements display a larger effect.

Table 2. Loci Associated with Multiple Elements and PC(s) in the Same Environment.

Environment	Chr	Pos (cM) †	Elements	PC(s)
NY05	1	400	Mn, Ni	PC11
NY05	3	323	Sr, Ca	--
NY05	5	201	Mn, Zn, P, S, Fe	PC1
NY06	1	532	Mn, Mg	--
IN09	4	306	Fe, K	--
IN10	2	213	Mo, Cd	PC2, PC4
NY12	5	203	Zn, Fe	PC7
FL05	1	230	B, Mn	--
FL05	4	159	Fe, Zn	--

†Average position of all element QTL, PC QTL are within 5 cM

Comparing PCs from different environments identified 52 PC pairs with similar loadings. Of these, 37 had no QTL for one or both of the PCs, consistent with a common environmental factor variable in those fields as the basis of that variation. Of the remaining 15 pairs, for which at least one QTL was detected for each member of the pair, five pairs had QTL that co-localized. In all five cases, the QTL overlapping between these pairs of PCs correspond to a large-effect single-element QTL. Six PC traits belonging to three correlated pairs, PC4 in NY05 and PC6 in IN09 ($r_p = 0.81$), PC4 in FL05 and PC3 in NY05 ($r_p = –0.84$), and PC3 in IN10 and PC2 in NC06 ($r_p = 0.89$), detected a QTL coinciding with a Mo QTL, a locus on chromosome 1 encoding the ortholog of the *A. thaliana* MOT1 molybdenum transporter. The same scenario exists for PC2 in IN09 and PC2 in NY05 ($r_p = –0.78$), both affected by the QTL on chromosome 2 that had a strong effect on Cd in our single-element QTL mapping experiments. Finally, PC8 in NC06 and PC5 in NY05 ($r_p = 0.76$) both map to a large-effect Ni QTL. Despite the resolution to QTL detected in a single-element analysis, in all of these cases correlations between loadings were not driven by a single element, but rather by similar loadings for most elements (Fig S2). In addition to overlaps at these strong-effect single-element QTL, 6 other pairs of correlated PCs have QTL

that do not overlap. Correlated PCs with QTL at different chromosomal positions in different environments could be due to states, such as iron deficiency, that may arise from distinct processes in each environment (e.g. soil pH or low Fe content) yet will generate a consistent physiological response. In these cases, the ionome displays similar trait covariance but different genetic architecture consistent with genotype by environment interactions.

The PC approach also detected a QTL that was found for different single elements depending on environment. The same region on chromosome 7 was identified as a QTL for three different elements in varying environments: Cu in NY05, NY12, IN10, and IN09, K in IN09, and Rb in NC06. In the mapping of QTL affecting the PC traits, we detected QTL at this position in some of the same environments as the single element QTL, NC06 and NY05, as well as in new environments, NY06 and SA10. In SA10, no QTL were mapped for Cu, Rb, or K alone. Yet, this locus was detected as significantly affecting variation in PC9 calculated from SA10, the loadings of which show a strong contribution from Cu and Rb. Likewise, in NY06, no QTL were mapped for Cu, Rb, or K, however, this locus was detected using PC6 in NY06 which has a strong loading contribution from K. No PC QTL were detected at the locus in NY12, IN09, or IN10. Thus, using PC traits in addition to single element traits can provide an improved estimate across different environments for the genetic effect on phenotypic variance for multi-element loci.

The identification of both unique and previously observed QTL through this multivariate approach demonstrates the complementary nature of working with trait covariance as well as the component traits and supports previous work showing that elemental traits are mechanistically interrelated. The repeated finding of results consistent with GxE led us to investigate this formally.

QTL by Environment Interactions

Our prior analyses found QTL by environment interactions contributing to accumulation of single elements [17]. Given element correlations and partially overlapping sets of element and PC QTL, we expect to detect QTL by environment interactions that impact multi-element traits. To look at the effects of environment on genetic regulation of multi-element phenotypes, we conducted another PCA, this time on element concentrations of lines from all environments combined. If the genetic and environmental variances do not interact, we expect some PCs will reflect environmental variance and others will reflect genetic variance. However, if the ionome is reporting on a summation of physiological status that results from genetic and environmental influences, some PCs calculated from ionomic traits should be both correlated with environmental factors and result in detectable QTL.

PCA across environments. The covariance between element accumulation data across all environments was summarized using principal components analysis. Elements prone to analytical artifacts (B, Na, Al, As) were removed prior to analysis. 16 across-environment PCs (aPCs) describing the covariation of the ionome were calculated for every RIL in every environment.

Out of a concern that the different lines present in each growout unduly influenced the construction of PCs specific to each environment, we performed the following tests. First, we looked at only those locations where two or more growouts were performed, so that location replication might be considered. Second, to identify a balanced sample set present in all environments, we identified the lines that were grown in all of these six growouts. PCA of the 16 element measurements was conducted across environments (Fig S3) and the loadings of each element into each PC were recorded. Thus, the loadings of the 16 elements in the PCA were

calculated from a set of common genotypic checks distributed within each environment. We used these loadings to calculate PCA projections (PJs) from all lines in all environments. In this way we made comparisons of the same calculated values in each environment. We found that the PJs and aPCs were strongly correlated; PJ1 and aPC1 were nearly identical (r_p = .998) and PJs 2–5 correlated with at least one of aPCs 2–5 at r_p > .66. The correlations between the loadings from PJs and aPCs reflected these same patterns. To reduce the incidence of artifacts or overfitting, aPCs accounting for less than 2% of the total variation were eliminated for further analyses, leaving seven aPCs.

Growth environment had a significant effect on all aPCs ($p < 0.001$). The first two aPCs were highly responsive to the environment (Fig 5). The lines from each environment cluster together when plotting aPC1 vs aPC2 values, with distinct separation between environments and years. In order to identify environmental factors responsible for ionome covariance, weather station and soil data from all environments except SA06 were recovered from databases (see methods). Correlations were calculated between season-long or quarter-length summaries of temperature and the aPC values for the nine environments. The weather variables, all temperature-based, were not correlated with aPCs in many cases, although correlations exceeding $r_p = 0.50$ were observed for aPCs 2,4, and 5 (Fig 6A). The strongest correlation observed for aPC1 was with average maximum temperature in the fourth quarter of the growing season ($r_p = 0.35$) (Fig 6B) while the highest observed for aPC2 was for average maximum temperature during the third quarter ($r_p = 0.58$) (Fig 6C). The relatively small number of environments, substantial non-independence of the weather variables, and likely contribution of factors other than temperature limit the descriptive power of these correlations.

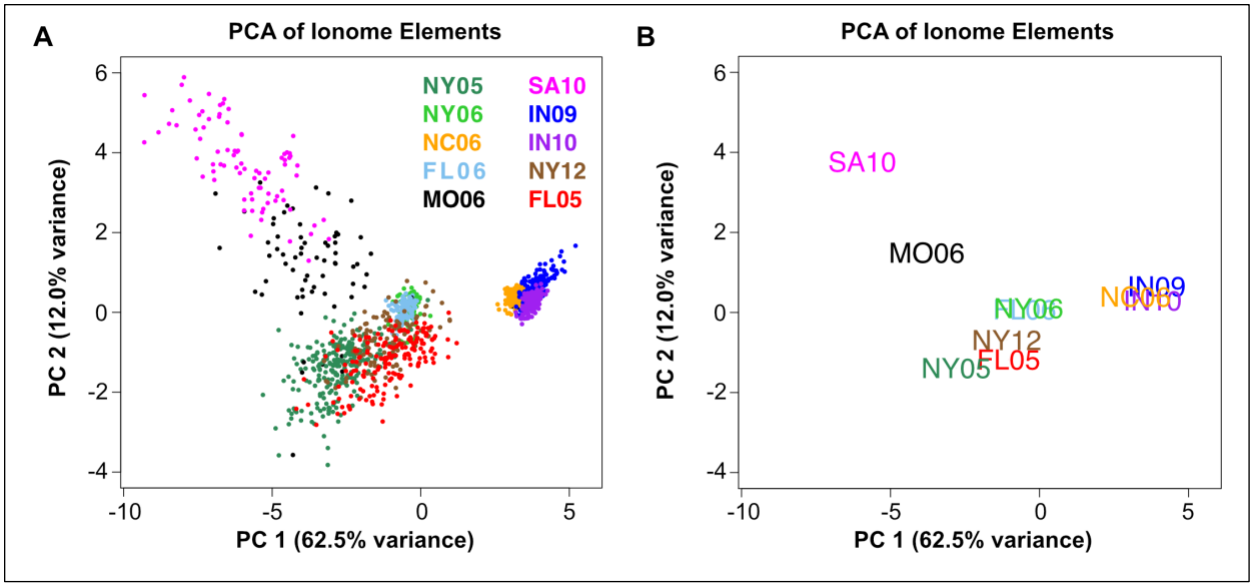


Fig 5. PCA Separates Lines by Environment. PC1 and PC2 separate lines by environment. Points correspond to lines, colored by their environment. (A) Across-environment PC1 vs PC2 values for each line, colored by environment. Percentage of total variance accounted for by each PC indicated on the axes. (B) Average across-environment PC1 vs PC2 values for all lines in each environment.

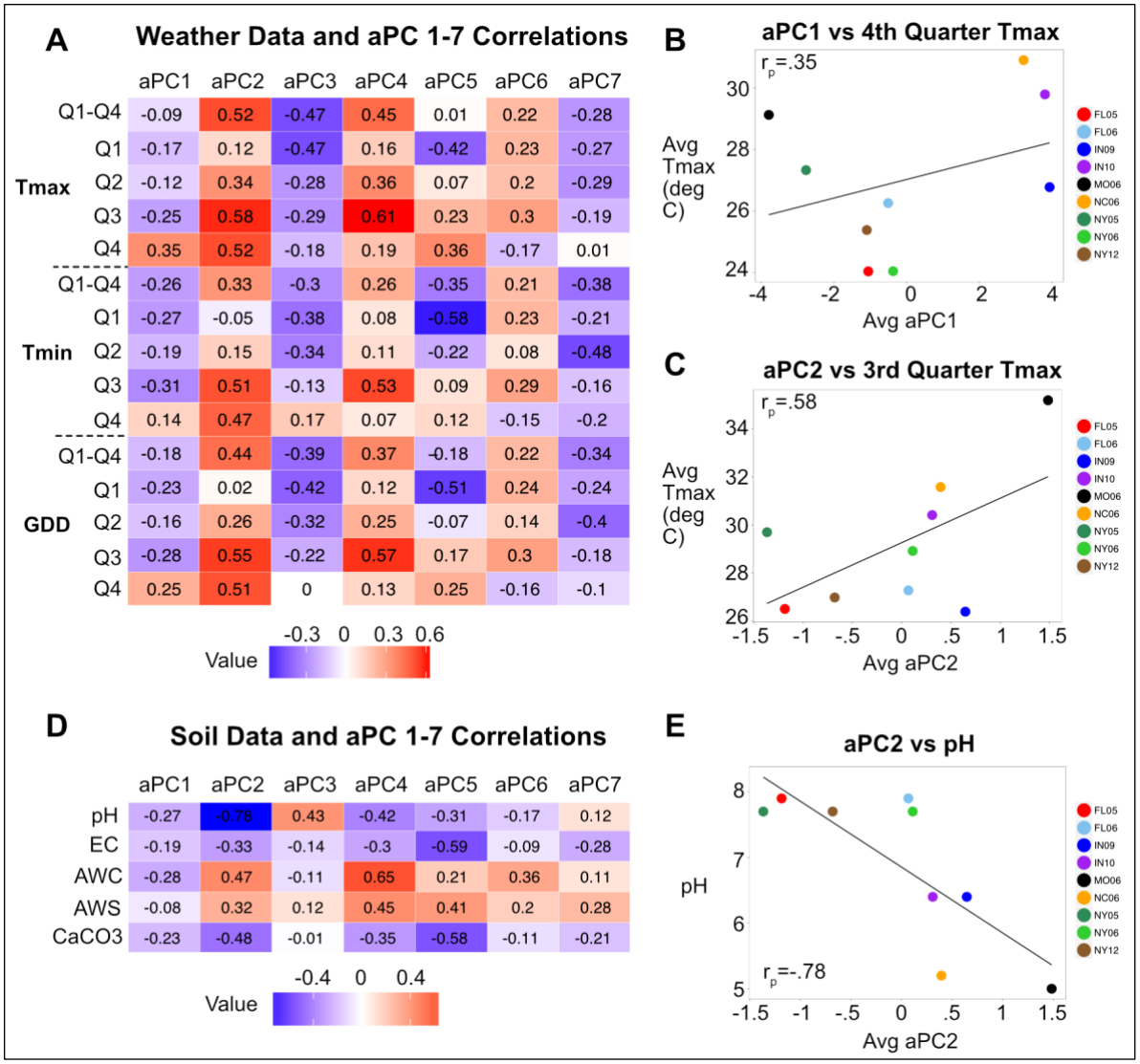


Fig 6. aPC and Weather Variable Correlations. (A) Heatmap showing Pearson correlation coefficients (r_p) between averaged aPC 1–7 values across environments and averages for maximum temperature, minimum temperature, and GDD across the growth season and for each quarter of the season. Red and blue intensities indicate strength of positive and negative correlations, respectively. (B) Average aPC1 values for 9 environments vs. average maximum temperature for each environment over the fourth quarter of the growing season. Points colored by environment. Pearson correlation coefficient is shown within the graph. (C) Average aPC2 values for nine environments vs. average maximum temperature for each environment over the 3rd quarter of the growing season. (D) Heatmap showing correlations between aPCs 1–7 and soil attributes: pH, electrical conductivity (EC), available water capacity (AWC), available water storage (AWS), and calcium carbonate (CaCO3). (E) Average aPC2 values vs. pH.

The lack of particularly strong correlations between the first two aPCs and temperature variables suggests that other variables, possibly field to field variation in soil composition, fertilizer application, humidity, or abiotic factors, are likely to have an influence. Correlations were also calculated between environment averages of the PCs and soil variables (Fig 6D). While the majority of these features were not found to be highly correlated with aPCs, we did observe a strong negative correlation between aPC2 and soil pH (r_p = –.78) (Fig 6E).

In order to determine genetic effects on these components, the calculated values for aPC1 through aPC7 were used as traits for QTL analysis in each of the 10 environments. Unlike the earlier described PCAs done in environments separately, these aPCs are calculated across all environments and are therefore comparable between environments. QTL mapping detected at least four loci controlling each aPC and a total of 38 QTL. Nine of these QTL were found in common across multiple environments and 29 were only detected in a single environment (Fig 7). Of the aPC QTL, the highest LOD score QTL were present in multiple environments and corresponded to the locations of the two strongest single element QTL previously detected from the same data (Mo on chromosome 1 and Cd on chromosome 2). The detection of QTL, together with the strong environmental determination of aPCs 1–7, demonstrates that ionomic covariation results from coordinate environmental and genetic variation.

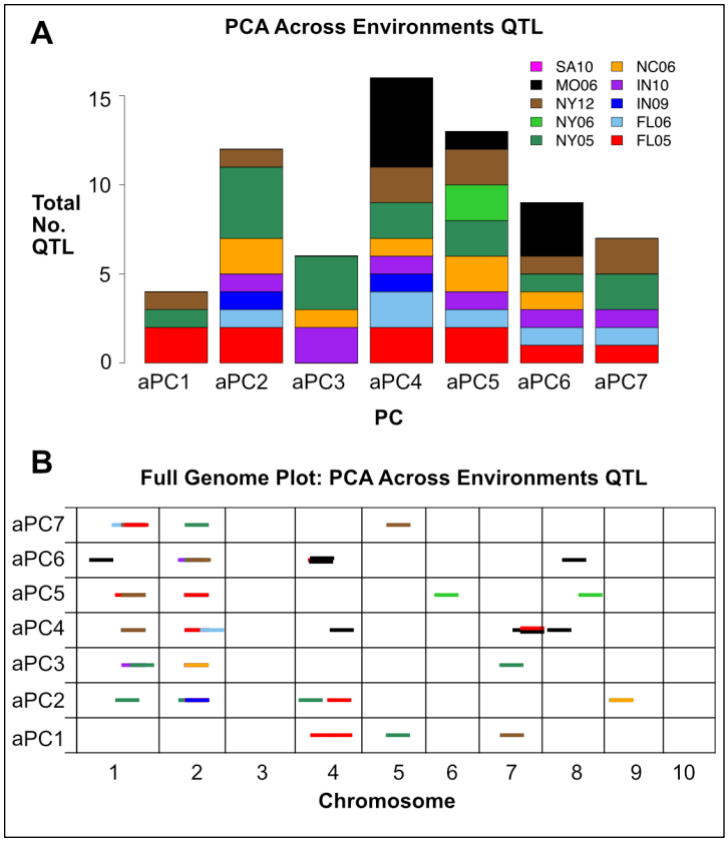


Fig 7. Across-Environment PCA QTL in 10 Environments. QTL identified for across environment PCA traits (aPCs 1–7). (A) Total number of QTL detected for each aPC, colored by environment. (B) Significant QTL ($\alpha = 0.05$) for aPCs 1–7. QTL location is shown across 10 chromosomes (in cM) on the x-axis. Dashes indicate QTL, with environment in which QTL was found designated by color. All dashes are the same length for visibility.

Based on the stochastic detection of QTL in only a subset of growth environments, substantial interaction between the environment aPC QTL is expected. A QTL of particular interest is the aPC2 QTL detected for Mo at the ortholog of the *MOT1* locus. Previous studies have demonstrated a connection between pH and molybdenum, with Mo availability in soil being increased by high pH. It was found that the *MOT1* locus in *A. thaliana* determines response to pH changes and resultant changes in Mo availability in an allele-specific manner, suggesting an adaptive role for variation in *MOT1* with respect to soil pH [18]. The correlation between aPC2 and pH was significant and aPC2 identified a QTL coinciding with a Mo QTL suggesting genetic

variation in pH-dependent changes to Mo availability across environments. The loading magnitude for Mo into aPC2 is 0.21 but Co, Ni, Rb, and Cd contribute even more, with loading magnitudes of 0.24, 0.46, 0.55, and 0.41, respectively. QTL for aPC2 also overlap with QTL for Cd and Ni. With aPC2 representing several elements, the correlation with soil pH and overlap with single element QTL may reflect a multi-element phenotype responding to changes in pH. Further investigation is needed to molecularly identify the genes underlying aPC QTL, their biological roles, and their interaction with specific environmental variables.

DISCUSSION

In this study, we demonstrate that multi-trait analysis is a valuable approach for understanding the ionome. The ionome is a homeostatic system, and effects on one element can affect other elements [1]. Many biological processes in maize have the potential to impact several elements. Indirect effects on a suite of elements have been demonstrated for numerous physiological states. Radial transport of nutrients is influenced in part by endodermal suberin, the structure and deposition of which can adapt in a highly plastic manner in response to deficiencies in K, S, Na, Fe, Zn, and Mn, potentially modifying transport of additional elements [19]. Other examples of indirect effects can be found in Arabidopsis *TSC10A* mutants with reduced 3-ketodihydrosphinganine (3-KDS) reductase activity. Because 3-KDS reductase is needed for synthesis of the sphingolipids that regulate ion transport through root membranes, these mutants exhibit a completely root-dependent leaf ionome phenotype of increased Na, K, and Rb, and decreased Mg, Ca, Fe, and Mo [20].

In line with the abundance of concerted element changes seen in ionome mutants, we detected elemental correlations and QTL that were present for more than one element.

Covariance observed in elements with similar orbital configurations, such as Ca and Sr or K and Rb, is expected due to related bonding properties and functions in redox reactions. The alkali metals K and Rb have been shown to display nearly identical absorption and distribution patterns [21]. Other elements are linked through co-regulation or common biological pathways. Phosphorous is a central nutrient in plant development and regulates other elements, complexing with cations in the form of phytic acid in maize seeds [22]. Phosphorous exhibited the greatest number of QTL overlap with other elements, including the cations K and Mg. Additional co-localized QTL included those between Zn and Fe, Mo and Mn, and the chemical analogs Ca and Sr. Zn and Fe can bind to the same metal transporters and metal-binding proteins and are thus reciprocally influenced in states of excess or deficiency [6, 23]. Three out of three of the Zn QTL that overlap with other elements involved overlap with Fe, demonstrating the genetic covariance of these elements. Mo and Mn have common roles in protein assimilation and nitrogen fixation [24, 25] and exhibit a regulatory relationship [26] which may explain their overlapping genetic features. The shared QTL detected in this study likely reflect genetic polymorphisms affecting the activity of multi-element regulatory genes or genetic changes targeted to a single element with pleiotropic effects on other elements due to homeostatic mechanisms or through concurrent multi-element behavior.

The 37 PC-specific loci identify loci in maize with the potential to expand our understanding of the genetic basis of ionome variation. The small population sizes used here limit our ability to interpret QTL-effect sizes, as overestimation of QTL effect, i.e. Beavis effect, is expected. Still, the large-effect QTL detected in our previous analysis [17] reappear as PC QTL. There is no reason to think that effect-size estimation will be any more accurate for PC than for single elements but careful simulations of correlated traits would be needed to

demonstrate this. Regardless, it seems likely that the loading of elements into PC will make these traits just as subject to effect size overestimation and may not provide additional support for the large size.

However, in the previous single element QTL analysis with this same dataset, we tested for overestimation using a more stringent permutation threshold and retained 31 of 63 location-specific QTL using a 99th percentile threshold. Biological mechanisms involving multi-element processes or synchronized element adjustments may drive the detection of unique PC QTL. For example, the ionome has been shown to exhibit tissue-dependent, multi-element changes in response to nitrogen availability [27]. A unique PC QTL could be detected at a nitrogen metabolism gene if variation at that gene confers additive effects on multiple elements. Variation in genes involved in adaptive responses to drought stress, soil nutrient deficiencies, or toxic micronutrient levels, can result in covariation among several elements without particularly strong effects on a single element [1, 6, 28], making such genes only identifiable as QTL when working with multivariate traits.

The majority of molecularly identified ionomic mutants have multi-element effects. In particular, mutants in genes involved in Casparian strip function and associated root-based element flow, including *MYB36* [29], *ESB1* [30], and *LOTR1* [31], all display pleiotropic effects on multiple element accumulation in the leaves. In some cases, QTL affecting these traits might be detected using both single and multi-element approaches, as was the case with the chromosome 5 QTL we mapped for P, S, Fe, Mn, and Zn, as well as for PC1. However, if the changes to a suite of elements are small for individual elements or uncontrolled environmental conditions inflate the magnitude of error in measuring the genetic effects, a multi-ionomic trait may be a better fit for QTL detection. The fact that we detect both overlapping and unique sets of

element and PC QTL suggests that single and multivariate approaches should be used in concert to avoid gaps in our understanding of element regulatory networks. The evidence suggests that some of the most interesting ionome homeostasis genes, including genes that are involved in environmental adaptation extending beyond the ionome, will be those best detected through multivariate methods.

In addition to being a tool for understanding the genetics of multi-element regulation, principal components also reflected environmental variation. An across-environment PCA of all lines was used to find variables that describe variation between lines among all 10 environments. The first two across-environment PCs capture most of the variation in the ionome across 10 different growouts, much of which is environmental. This can be seen in the ability of aPC1 and aPC2 to separate growouts by location and, in some cases, different years within a location. Thus, components from a PCA done across environments can capture the impact of environment on the ionome as a whole.

In our across-environment analysis, to account for different sets of IBM lines within environments, we tested an approach of projecting loadings from a PCA on a smaller set of lines onto the full data set. The similarity of the PJs and aPCs led us to conclude that the sampling effects of having different subsets of lines in each environment had little effect on the trait covariance estimation. This approach to validate aPCs may be useful in other studies that seek to connect data from disparate experiments and federate data collected by multiple laboratories. The method of deriving traits across environments using a small set of genotypic checks opens up the possibility of using multi-trait correlations across environments to permit very large scale GxE mapping experiments on data sets not initially intended for this purpose. Retrospective analysis of data, or further data generation from preexisting biological material present in both

public and private spheres, is enabled by this approach. For example, multiple association panels have been constructed for trait mapping in maize. Typically, comparison of multi-trait correlations across different populations is inhibited by our inability to ensure the 1:1 correspondence of traits. By using the subset of lines common to all mapping populations to create a projection, comparable traits could be reflected onto to full datasets for comprehensive genetic evaluation and the loci detected in each panel could then be compared, as we have done here.

PCA on all environments is a way to find variation resulting from environmental factors that impact multiple elements, for example weather or soil variables. The weather data available to us for this study was limited to maximum and minimum temperature. Temperature can alter element accumulation by influencing transpiration rate which in turn modulates elemental movement [30, 32, 33]. We observed the strongest correlations for aPC1 and aPC2 during the third and fourth quarters of the growing season. Because seed filling occurs in the latter part of the season, temperature during this time could have a pronounced effect on seed elemental composition. However, the lack of striking correlations between environmental components and the projections and aPCs suggests environmental factors other than temperature must be the strongest factors. Information on soil properties provided insight into a potential driver of the environmental variability captured by aPC2, with a strong negative correlation between aPC2 and soil pH. Soil pH alters element availability in soil, and pH differences between locations should result in different kernel ionomes. Although soil element content measurements were not available for this dataset, differences in soil element concentration could also impact element covariation.

QTL were mapped to the aPCs that describe whole ionome variation across environments. These loci may encompass genes that pleiotropically affect the ionome in an environmentally-responsive manner. The correlation between aPC2 with pH as well as the finding of an aPC2 QTL for Mo exemplifies the possibility of using across-environment PCA to detect element homeostasis loci that respond to a particular environmental or soil variable and produce a multi-element phenotype. To the extent that these differences are adaptive, these alleles can contribute to local adaptation to soil environment and nutrient availability. The identification of aPC QTL indicates that the variation captured by aPCs has both environmental and genetic components. Our previous study using single element traits found extensive GxE in this dataset through formal tests, so it is not surprising that we see a large environmental component as well as genetic factors contributing to variation in the across-environment PCs. Experiments with more extensive weather and soil data, or carefully manipulated environmental contrasts, are needed to create models with additional covariates and precisely represent environmental impacts. Considering location and geographical information, such as proximity to industrial sites or distance from the ocean, might add to the predictive ability of such models. This multivariate approach could be especially powerful in studies with extensive and consistent environmental variable recording, such as the "Genomes to Fields" Initiative, where specific environmental variables could be included in QTL models of multi-element GxE.

ACKNOWLEDGEMENTS

The authors would especially like to thank our field collaborators Sherry Flint-Garcia, Peter Balint-Kurti, Torbert Rocheford, Jonathan Lynch, and Robert Snyder for their dedicated efforts to provide the seeds analyzed for this project. This work was supported by funding from the National Science Foundation (IOS-1126950, IOS-1450341), the USDA Agricultural Research Service (5070-21000-039-00D). AA was a recipient of a Danforth Plant Science Fellowship from the Donald Danforth Plant Science Center. The authors declare no conflict of interest.

REFERENCES

1. Baxter IR, Vitek O, Lahner B, Muthukumar B, Borghi M, Morrissey J et al. The leaf ionome as a multivariable system to detect a plant's physiological status. Proceedings of the National Academy of Sciences. 2008;105: 12081-12086.
2. Korshunova YO, Eide D, Clark WG, Guerinot ML, Pakrasi HB. The IRT1 protein from Arabidopsis thaliana is a metal transporter with a broad substrate range. Plant molecular biology. 1999;40: 37-44.
3. Broadley MR, Hammond JP, King GJ, Astley D, Bowen HC, Meacham MC et al. Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in Brassica oleracea. Plant Physiol. 2008;146: 1707-1720.
4. Buescher E, Achberger T, Amusan I, Giannini A, Ochsenfeld C, Rus A et al. Natural genetic variation in selected populations of Arabidopsis thaliana is associated with ionomic differences. PLoS One. 2010;5: e11081.
5. Baxter IR, Gustin JL, Settles AM, Hoekenga OA. Ionomic characterization of maize kernels in the intermated B73√ó Mo17 population. Crop Science. 2013;53: 208-220.
6. Baxter I. Ionomics: studying the social network of mineral nutrients. Current Opinion in Plant Biology. 2009;12: 381-386.
7. Burton AL, Johnson J, Foerster J, Hanlon MT, Kaeppler SM, Lynch JP et al. QTL mapping and phenotypic variation of root anatomical traits in maize (Zea mays L.). Theoretical and Applied Genetics. 2015;128: 93-106.
8. Bouchet S, Bertin P, Presterl T, Jamin P, Coubriche D, Gouesnard B et al. Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits. Heredity. 2017;118: 249-259.
9. Frey FP, Presterl T, Lecoq P, Orlik A, Stich B. First steps to understand heat tolerance of

temperate maize at adult stage: identification of QTL across multiple environments with connected segregating populations. Theoretical and Applied Genetics. 2016;129: 945-961.
10. Buescher EM, Moon J, Runkel A, Hake S, Dilkes BP. Natural Variation at sympathy for the ligule Controls Penetrance of the Semidominant Liguleless narrow-R Mutation in Zea mays. G3: Genes Genomes Genetics. 2014;4: 2297-2306.
11. Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R, Connolly EL et al. Integration of experiments across diverse environments identifies the genetic determinants of variation in Sorghum bicolor seed element composition. Plant physiology. 2016; pp-01971.
12. Topp CN, Iyer-Pascuzzi AS, Anderson JT, Lee C-R, Zurek PR, Symonova O et al. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proceedings of the National Academy of Sciences. 2013;110: E1695-E1704.
13. Liu Z, Garcia A, McMullen MD, Flint-Garcia SA. Genetic Analysis of Kernel Traits in Maize-Teosinte Introgression Populations. G3: Genes Genomes Genetics. 2016;6: 2523.
14. Choe E, Rocheford TR. Genetic and QTL analysis of pericarp thickness and ear architecture traits of Korean waxy corn germplasm. Euphytica. 2012;183: 243-260.
15. Zhang N, Gibon Y, Gur A, Chen C, Lepak N, Höhne M et al. Fine Quantitative Trait Loci Mapping of Carbon and Nitrogen Metabolism Enzyme Activities and Seedling Biomass in the Intermated Maize IBM Mapping Population. Plant Physiology. 2010.
16. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D et al. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant molecular biology. 2002;48: 453-461.
17. Asaro A, Ziegler G, Ziyomo C, Hoekenga OA, Dilkes BP, Baxter I. The Interaction of Genotype and Environment Determines Variation in the Maize Kernel Ionome. G3: Genes Genomes Genetics. 2016;6: 4175-4183.
18. Poormohammad Kiani S, Trontin C, Andreatta M, Simon M, Robert T, Salt DE et al. Allelic Heterogeneity and Trade-Off Shape Natural Variation for Response to Soil Micronutrient. PLOS Genetics. 2012;8: e1002814.
19. Barberon M, Vermeer J, De Bellis D, Wang P, Naseer S, Andersen T et al. Adaptation of Root Function by Nutrient-Induced Plasticity of Endodermal Differentiation. Cell. 2016;164: 447-459.
20. Chao DY, Gable K, Chen M, Baxter I, Dietrich CR, Cahoon EB et al. Sphingolipids in the Root Play an Important Role in Regulating the Leaf Ionome in Arabidopsis thaliana. Plant Cell. 2011;23: 1061-1081.
21. Läuchli A, Epstein E. Transport of potassium and rubidium in plant roots: the significance of calcium. Plant Physiology. 1970;45: 639.
22. López-Arredondo DL, Leyva-González MA, González-Morales SI, López-Bucio J, Herrera-Estrella L. Phosphate nutrition: improving low-phosphate tolerance in crops. Annual Review of Plant Biology. 2014;65: 95-123.
23. Lin Y-F, Liang H-M, Yang S-Y, Boch A, Clemens S, Chen C-C et al. Arabidopsis IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. New Phytologist. 2009;182: 392-404.
24. Mulder EG. Importance of molybdenum in the nitrogen metabolism of microorganisms and higher plants. Plant and Soil. 1948;1: 94-119.
25. Mulder EG, Gerretsen FC. Soil manganese in relation to plant growth. Adv Agron. 1952;4: 221-277.

26. Millikan CR. Antagonism between molybdenum and certain heavy metals in plant nutrition. Nature. 1948;161: 528.
27. Chu Q, Watanabe T, Shinano T, Nakamura T, Oka N, Osaki M et al. The dynamic state of the ionome in roots, nodules, and shoots of soybean under different nitrogen status and at different growth stages. 2016:488-498.
28. Baxter I, Dilkes BP. Elemental Profiles Reflect Plant Adaptations to the Environment. Science. 2012;336: 1661-1663.
29. Kamiya T, Borghi M, Wang P, Danku JMC, Kalmbach L, Hosmani PS et al. The MYB36 transcription factor orchestrates Casparian strip formation. Proceedings of the National Academy of Sciences. 2015;112: 10533-10538.
30. Baxter I, Hosmani PS, Rus A, Lahner B, Borevitz JO, Muthukumar B et al. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. PLoS Genet. 2009;5: e1000492.
31. Li B, Kamiya T, Kalmbach L, Yamagami M, Yamaguchi K, Shigenobu S et al. Role of LOTR1 in Nutrient Transport through Organization of Spatial Distribution of Root Endodermal Barriers. Current Biology. 2017;27: 758-765.
32. Barber SA, Mackay AD, Kuchenbuch RO, Barraclough PB. Effects of soil temperature and water on maize root growth. Plant and Soil. 1988;111: 267-269.
33. Mozafar A, Schreiber P, Oertli JJ. Photoperiod and root-zone temperature: Interacting effects on growth and mineral nutrients of maize. Plant and Soil. 1993;153: 71-78.
34. Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64: 641-656.

SUPPORTING INFORMATION

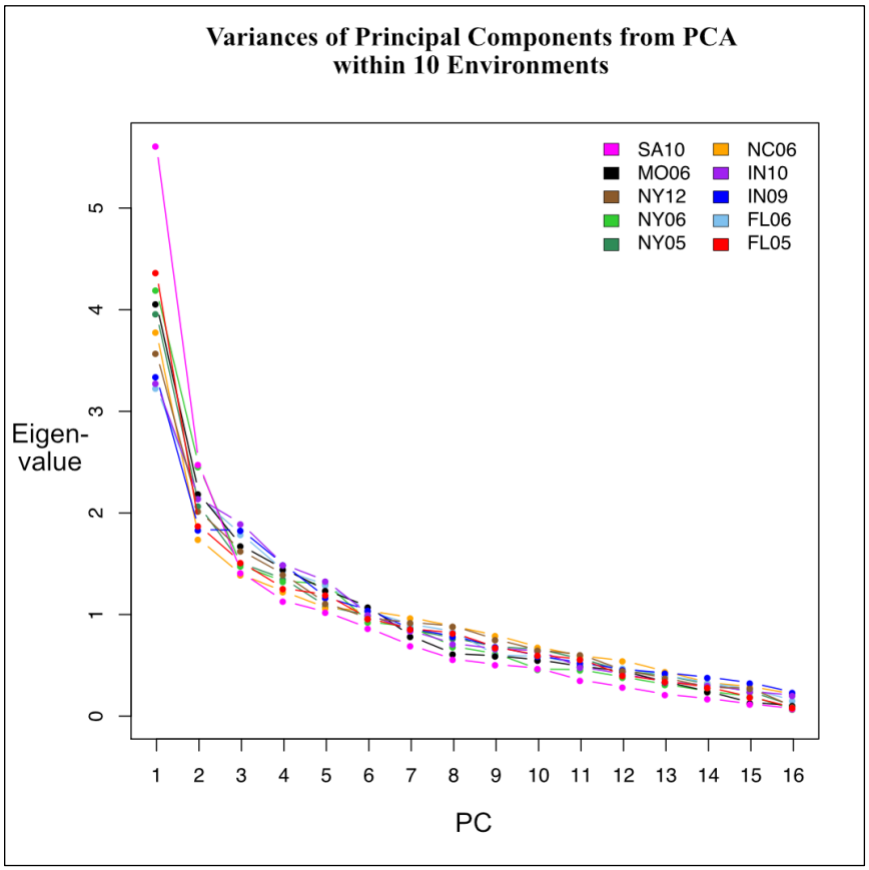


Fig S1. Variances of Principal Components from PCA within 10 Environments. Eigenvalues (amount of variation explained) for each PC are shown on the y-axis. Lines are colored by environment.

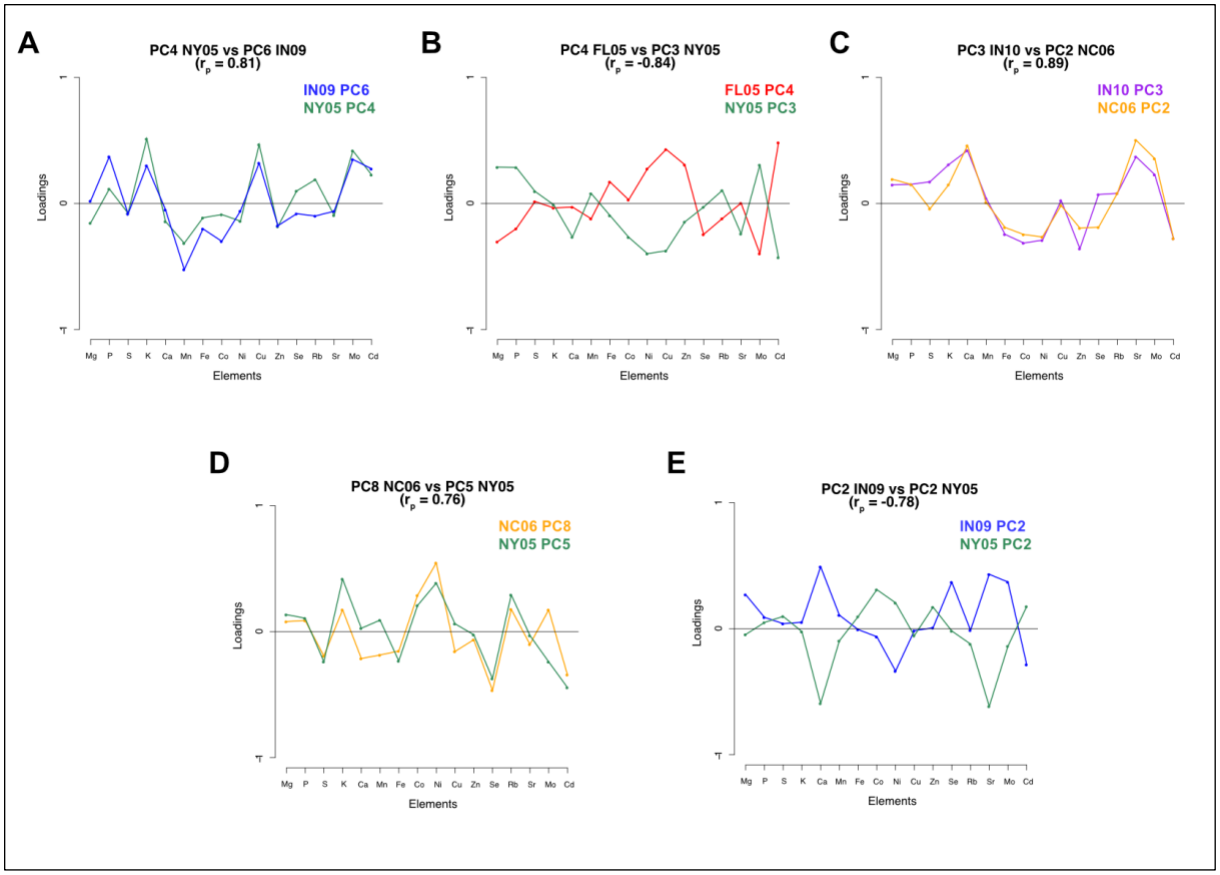


Fig S2. Loadings of Principal Components from Different Environments. Loadings for each element are plotted for PCs from different environments. Loadings of PCs plotted on the same graph are correlated as indicated. PCs shown in (A), (B), and (C) all have a QTL coinciding with Mo QTL on chromosome 1. PCs shown in (D) have a QTL coinciding with Cd QTL on chromosome 2. PCs shown in (E) have a QTL coinciding with Ni QTL on chromosome 9.

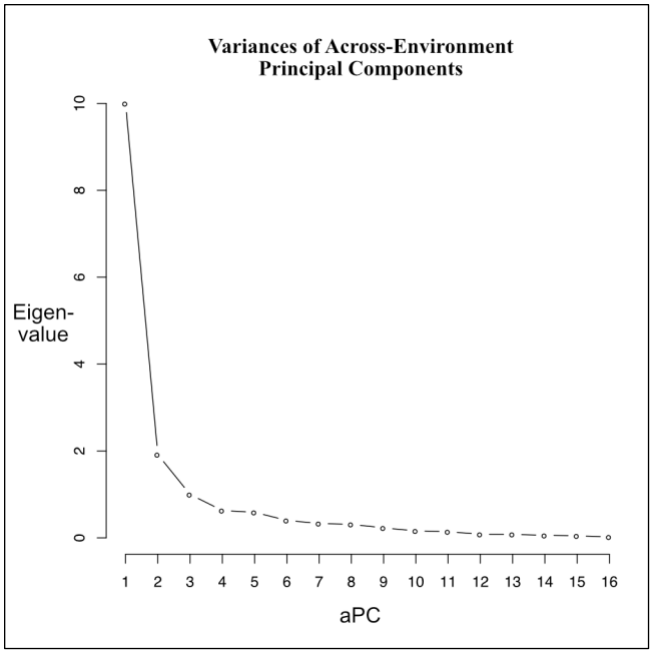


Fig S3. Variances of Principal Components from PCA on Lines from all Environments. Eigenvalues (amount of variation explained) for each aPC are shown on the y-axis.

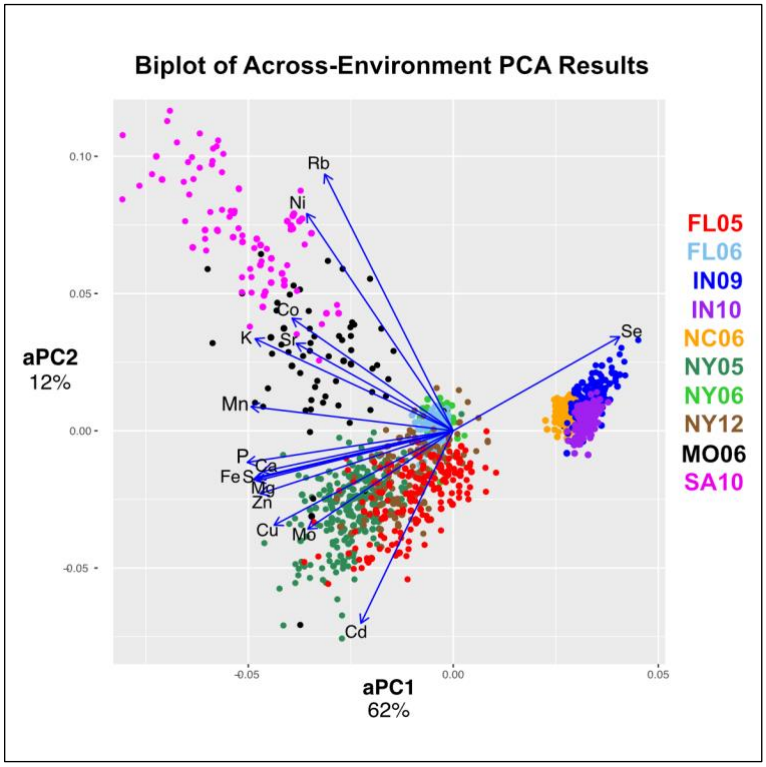


Fig S4. aPC1 and aPC2 Loadings Biplot. PCA plots showing aPC1 and aPC2 loadings. Variance explained for each PC is indicated along axes.

Table S1. Within Environment PCs QTL Counts.

	Overall Total	FL05	FL06	IN09	IN10	MO06	NC06	NY05	NY06	NY12	SA10
PC1	5	2	0	0	0	1	1	1	0	0	0
PC2	6	0	0	1	0	0	1	4	0	0	0
PC3	10	1	0	0	1	0	0	5	1	2	0
PC4	10	2	0	1	0	0	1	5	0	1	0
PC5	22	1	1	0	1	0	1	6	9	1	2
PC6	9	0	0	1	1	0	2	3	1	1	0
PC7	7	0	0	2	1	0	1	0	0	3	0
PC8	3	0	0	1	1	0	1	0	0	0	0
PC9	6	0	0	1	0	0	1	1	0	1	2
PC10	4	0	0	1	0	0	1	2	0	0	0
PC11	2	0	0	0	0	0	0	2	0	0	0
PC12	6	0	0	2	1	0	3	0	0	0	0
PC13	2	0	0	1	0	0	0	1	0	0	0
PC14	2	0	0	0	1	0	0	0	0	0	1
PC15	4	0	0	0	0	0	1	2	0	0	1
PC16	2	1	0	0	0	0	0	1	0	0	0

Table S2. PC Loadings and Element QTL Overlap.

Each PC QTL is shown with the QTL location, environment, and LOD score information. Elements with the top 5 loadings into the PC trait are listed. Unique PC QTL are highlighted (QTL with no overlap within 25 cM). For PC QTL that have element QTL within 10 cM, the element traits for these QTL are listed.

Trait	Chr	Pos	Envirs	MaxLOD	MaxPerm	Elements with top 5 PC loadings (increasing order)	Elemental QTL within 25 cM?	Elements with QTL within 10 cM
PC1	4	173.9	FL05	5.11	3.69	Mn, Fe, S, Mg, P	YES	Mn, K
PC1	**4**	**287.6**	**FL05**	**5.09**	**3.69**	**Mn, Fe, S, Mg, P**	**NO**	
PC3	1	404.0	FL05	3.93	3.71	Mg, Fe, Mn, K, Rb	YES	Mn
PC4	1	252.4	FL05	4.33	3.73	Mg, Zn, Mo, Cu, Cd	YES	
PC4	1	380.6	FL05	11.70	3.73	Mg, Zn, Mo, Cu, Cd	YES	
PC5	2	215.0	FL05	6.35	3.68	Sr, Cu, S, Cd, Co	YES	Cd
PC5	1	378.0	FL06	4.24	3.68	Fe, P, Mn, Cd, Mo	YES	Mo
PC2	2	216.9	IN09	5.92	3.61	Ni, Se, Mo, Sr, Ca	YES	Cd
PC4	2	203.2	IN09	5.08	3.63	Mg, Cd, Mo, Se, Co	YES	
PC6	1	378.0	IN09	7.14	3.56	K, Cu, Mo, P, Mn	YES	Mo
PC7	**3**	**358.5**	**IN09**	**3.91**	**3.63**	**Cd, S, K, Mn, Mg**	**NO**	
PC7	4	300.0	IN09	3.72	3.63	Cd, S, K, Mn, Mg	YES	Fe, K
PC8	9	7.7	IN09	8.87	3.62	P, K, Cd, Se, Ni	YES	Ni
PC9	**9**	**302.2**	**IN09**	**3.78**	**3.74**	**Co, Mo, Rb, Fe, S**	**NO**	
PC10	2	236.7	IN09	3.78	3.56	Co, Cu, Cd, Mo, S	YES	Ni
PC12	**1**	**136.5**	**IN09**	**4.08**	**3.62**	**Co, Cu, Zn, Fe, P**	**NO**	
PC12	**3**	**267.9**	**IN09**	**4.06**	**3.62**	**Co, Cu, Zn, Fe, P**	**NO**	
PC13	**5**	**33.0**	**IN09**	**4.75**	**3.77**	**K, Rb, Zn, Se, Cu**	**NO**	
PC3	1	378.0	IN10	3.60	3.57	K, Co, Zn, Sr, Ca	YES	Mo
PC5	2	211.7	IN10	5.36	3.65	Ni, Fe, Mn, Sr, Mo	YES	Mo, Cd
PC6	2	209.5	IN10	4.48	3.68	S, Ni, Mn, Cd, Rb	YES	Mo, Cd
PC7	4	315.8	IN10	4.45	3.75	Zn, Rb, K, Mg, Cd	YES	
PC8	1	377.3	IN10	7.62	3.71	Sr, Co, S, Ni, Mo	YES	Mo
PC12	**2**	**102.2**	**IN10**	**4.32**	**3.65**	**Zn, Mn, Cu, Co, Se**	**NO**	
PC1	10	95.5	MO06	3.79	3.70	S, Rb, Fe, K, P	YES	Rb
PC1	7	167.0	NC06	4.62	3.65	Rb, Mg, Zn, P, K	YES	Rb
PC2	1	378.0	NC06	4.47	3.70	Ni, Cd, Mo, Ca, Sr	YES	Mo
PC4	9	16.8	NC06	4.36	3.58	Mn, P, Ca, Co, Se	YES	Ni
PC5	**3**	**358.7**	**NC06**	**4.48**	**3.50**	**K, Mn, Cu, Rb, Mg**	**NO**	
PC6	**1**	**244.9**	**NC06**	**4.92**	**3.62**	**S, Ni, Mn, Fe, Mo**	**NO**	
PC6	2	217.9	NC06	4.11	3.62	S, Ni, Mn, Fe, Mo	YES	Cd
PC7	2	215.0	NC06	12.03	3.75	Fe, S, Ni, Cu, Cd	YES	Cd
PC8	9	8.9	NC06	6.85	3.71	Ca, Co, Cd, Se, Ni	YES	Ni
PC9	**3**	**148.6**	**NC06**	**4.39**	**3.59**	**Se, Ni, Cu, Zn, Fe**	**NO**	
PC10	**3**	**156.8**	**NC06**	**3.91**	**3.62**	**Sr, S, Cu, Rb, Mn**	**NO**	
PC12	**1**	**113.8**	**NC06**	**8.30**	**3.75**	**Mn, Rb, Fe, Cd, S**	**NO**	
PC12	**1**	**515.3**	**NC06**	**4.74**	**3.75**	**Mn, Rb, Fe, Cd, S**	**NO**	
PC12	**9**	**146.3**	**NC06**	**4.32**	**3.75**	**Mn, Rb, Fe, Cd, S**	**NO**	
PC1	5	203.8	NY05	6.92	3.73	Mn, Zn, Fe, Mg, P	YES	Mn, Fe, Zn, P, S
PC2	2	216.9	NY05	5.57	3.61	Cd, Ni, Co, Ca, Sr	YES	Cd
PC2	3	331.0	NY05	5.48	3.61	Cd, Ni, Co, Ca, Sr	YES	Sr, Ca
PC2	7	193.8	NY05	4.63	3.61	Cd, Ni, Co, Ca, Sr	YES	Sr

PC2	9	0.9	NY05	6.34	3.61	Cd, Ni, Co, Ca, Sr	YES	Ni
PC3	1	377.3	NY05	14.78	3.78	Mg, Mo, Cu, Ni, Cd	YES	Mo
PC3	2	211.0	NY05	12.89	3.78	Mg, Mo, Cu, Ni, Cd	YES	Cd
PC3	9	5.4	NY05	5.66	3.78	Mg, Mo, Cu, Ni, Cd	YES	Ni
PC3	**10**	**87.8**	**NY05**	**4.63**	**3.78**	**Mg, Mo, Cu, Ni, Cd**	**NO**	
PC3	**10**	**121.6**	**NY05**	**7.30**	**3.78**	**Mg, Mo, Cu, Ni, Cd**	**NO**	
PC4	1	378.0	NY05	8.49	3.55	Cd, Mn, Mo, Cu, K	YES	Mo
PC4	2	218.3	NY05	3.98	3.55	Cd, Mn, Mo, Cu, K	YES	Cd
PC4	3	325.8	NY05	4.18	3.55	Cd, Mn, Mo, Cu, K	YES	Sr, Ca
PC4	4	178.9	NY05	4.08	3.55	Cd, Mn, Mo, Cu, K	YES	K
PC4	7	165.9	NY05	7.54	3.55	Cd, Mn, Mo, Cu, K	YES	Cu
PC5	1	171.4	NY05	5.20	3.68	Rb, Se, Ni, K, Cd	YES	K
PC5	2	208.9	NY05	7.62	3.68	Rb, Se, Ni, K, Cd	YES	Cd
PC5	4	374.9	NY05	6.67	3.68	Rb, Se, Ni, K, Cd	YES	
PC5	7	150.7	NY05	3.96	3.68	Rb, Se, Ni, K, Cd	YES	K
PC5	9	7.7	NY05	8.48	3.68	Rb, Se, Ni, K, Cd	YES	Ni
PC5	9	136.6	NY05	4.83	3.68	Rb, Se, Ni, K, Cd	YES	Ni
PC6	1	378.0	NY05	14.86	3.58	K, Cd, Rb, Ni, Mo	YES	Mo
PC6	2	214.6	NY05	5.77	3.58	K, Cd, Rb, Ni, Mo	YES	Cd
PC6	9	8.3	NY05	5.00	3.58	K, Cd, Rb, Ni, Mo	YES	Ni
PC9	9	5.4	NY05	4.09	3.72	Cd, Zn, Ni, Co, Se	YES	Ni
PC10	1	385.7	NY05	4.33	3.60	Mn, Zn, Mo, Cd, Fe	YES	Mo
PC10	**10**	**147.6**	**NY05**	**4.12**	**3.60**	**Mn, Zn, Mo, Cd, Fe**	**NO**	
PC11	**6**	**128.6**	**NY05**	**4.69**	**3.72**	**Mg, Mn, K, Cu, S**	**NO**	
PC11	6	256.4	NY05	4.73	3.72	Mg, Mn, K, Cu, S	YES	
PC13	1	232.0	NY05	4.07	3.63	K, Mo, P, Mg, Mn	YES	Mn
PC3	**6**	**42.5**	**NY06**	**3.64**	**3.60**	**Cd, Cu, Rb, Ni, Co**	**NO**	
PC5	**1**	**167.0**	**NY06**	**3.71**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**1**	**169.7**	**NY06**	**14.43**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**1**	**271.2**	**NY06**	**21.58**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	1	379.7	NY06	19.95	3.03	Mn, Mg, Co, Se, Mo	YES	Mo
PC5	2	98.3	NY06	14.33	3.03	Mn, Mg, Co, Se, Mo	YES	Mo
PC5	**2**	**257.5**	**NY06**	**10.76**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**4**	**75.9**	**NY06**	**4.43**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**6**	**109.3**	**NY06**	**8.99**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**6**	**158.0**	**NY06**	**16.68**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC5	**8**	**355.7**	**NY06**	**6.57**	**3.03**	**Mn, Mg, Co, Se, Mo**	**NO**	
PC6	**7**	**162.3**	**NY06**	**4.33**	**3.66**	**Mo, Sr, Mn, K, Cd**	**NO**	
PC3	2	214.1	NY12	5.48	3.65	S, Rb, Ni, K, Cd	YES	Cd
PC3	9	0.0	NY12	3.68	3.65	S, Rb, Ni, K, Cd	YES	Ni
PC4	**3**	**221.5**	**NY12**	**3.72**	**3.70**	**Mn, Se, Co, K, Cu**	**NO**	
PC5	**5**	**150.9**	**NY12**	**3.59**	**3.56**	**Fe, Se, Mo, K, Ni**	**NO**	
PC6	2	210.8	NY12	3.72	3.59	Mg, Zn, Cd, Mo, Se	YES	Cd
PC7	2	242.5	NY12	4.62	3.58	Rb, Cd, S, Mg, Co	YES	Ni
PC7	**5**	**107.0**	**NY12**	**4.72**	**3.58**	**Rb, Cd, S, Mg, Co**	**NO**	
PC7	**6**	**255.4**	**NY12**	**4.07**	**3.58**	**Rb, Cd, S, Mg, Co**	**NO**	
PC9	**1**	**342.2**	**NY12**	**5.12**	**3.69**	**Zn, Mo, Se, Rb, Ni**	**NO**	
PC5	**1**	**83.5**	**SA10**	**4.65**	**3.68**	**Co, Cd, Mg, K, Rb**	**NO**	
PC5	**4**	**382.9**	**SA10**	**3.83**	**3.68**	**Co, Cd, Mg, K, Rb**	**NO**	
PC9	**1**	**418.2**	**SA10**	**4.77**	**3.28**	**Fe, Ni, Co, Rb, Cu**	**NO**	
PC9	**7**	**169.8**	**SA10**	**4.25**	**3.28**	**Fe, Ni, Co, Rb, Cu**	**NO**	

Table S3. PC Variance Proportions and Loadings Across 10 Environments.

FL05	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	2.09	1.37	1.23	1.12	1.09	0.98	0.93	0.91	0.82	0.78	0.75	0.64	0.59	0.53	0.44	0.29
Proportion of Variance	0.27	0.12	0.09	0.08	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.01	0.01
Cumulative Proportion	0.27	0.39	0.48	0.56	0.64	0.70	0.75	0.80	0.85	0.88	0.92	0.94	0.96	0.98	0.99	1.00
Mg	0.33	-0.03	0.20	-0.30	0.09	-0.24	0.07	0.13	-0.42	0.11	0.08	0.41	-0.15	-0.18	0.50	-0.11
P	0.38	-0.13	-0.04	-0.20	0.11	0.00	0.12	0.26	-0.27	0.20	-0.02	0.13	0.18	0.21	-0.69	0.14
S	0.33	0.01	0.12	0.02	0.36	-0.15	0.05	-0.10	0.15	0.24	0.40	-0.57	-0.37	0.04	0.02	0.08
K	0.24	0.01	-0.56	-0.04	-0.05	-0.20	0.32	-0.13	-0.01	-0.05	-0.14	-0.17	0.28	0.47	0.34	-0.03
Ca	0.26	0.53	0.12	-0.03	-0.20	-0.04	-0.07	-0.11	0.10	-0.11	0.09	0.05	-0.03	0.13	-0.21	-0.69
Mn	0.28	-0.05	0.33	-0.12	-0.07	0.22	0.26	-0.11	-0.19	-0.50	-0.38	-0.41	0.07	-0.23	0.00	0.06
Fe	0.32	-0.18	0.29	0.17	0.02	0.08	-0.01	0.00	0.43	0.20	0.13	0.09	0.66	-0.16	0.16	-0.03
Co	-0.05	-0.33	0.00	0.03	-0.63	0.01	0.17	-0.33	-0.28	0.07	0.50	-0.08	0.02	-0.07	-0.09	-0.03
Ni	0.17	-0.17	-0.09	0.28	-0.17	-0.54	-0.44	0.39	-0.05	-0.39	0.07	-0.15	0.05	-0.08	-0.04	0.02
Cu	0.23	0.10	-0.15	0.43	-0.30	0.21	-0.07	0.19	-0.18	0.52	-0.37	-0.17	-0.17	-0.20	0.09	-0.06
Zn	0.26	-0.31	0.05	0.31	-0.11	0.13	0.30	0.09	0.37	-0.24	-0.03	0.38	-0.47	0.22	-0.01	0.02
Se	0.16	-0.32	0.08	-0.25	-0.09	-0.16	-0.49	-0.53	0.13	0.16	-0.39	0.05	-0.13	0.15	-0.05	0.00
Rb	0.23	-0.01	-0.59	-0.12	0.14	0.05	0.02	-0.20	0.17	-0.13	0.07	0.14	-0.06	-0.65	-0.15	-0.05
Sr	0.23	0.56	0.06	0.00	-0.28	-0.07	-0.07	-0.16	0.08	-0.05	0.08	0.17	-0.01	0.02	0.02	0.69
Mo	0.16	-0.07	-0.17	-0.40	-0.16	0.59	-0.39	0.32	0.08	-0.10	0.23	-0.12	-0.04	0.18	0.21	0.02
Cd	0.15	0.00	-0.05	0.48	0.38	0.31	-0.28	-0.34	-0.43	-0.19	0.16	0.15	0.09	0.17	0.04	0.00

FL06	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.80	1.47	1.34	1.20	1.14	1.02	0.95	0.91	0.77	0.77	0.72	0.69	0.63	0.57	0.52	0.40
Proportion of Variance	0.20	0.13	0.11	0.09	0.08	0.06	0.06	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.01
Cumulative Proportion	0.20	0.34	0.45	0.54	0.62	0.68	0.74	0.79	0.83	0.87	0.90	0.93	0.95	0.97	0.99	1.00
Mg	0.36	-0.23	0.23	-0.05	0.04	-0.34	0.05	-0.03	-0.12	0.41	0.14	-0.15	-0.02	0.25	-0.55	-0.23
P	0.42	-0.04	0.10	-0.18	0.23	-0.10	-0.21	-0.22	-0.06	0.20	-0.08	-0.38	0.23	0.04	0.58	0.21
S	0.42	-0.11	0.09	-0.10	0.11	0.04	0.16	0.22	0.30	0.22	0.00	0.38	-0.40	-0.50	0.11	0.08
K	0.32	0.24	0.08	-0.18	0.02	0.47	-0.07	0.01	0.23	-0.25	0.53	-0.09	-0.01	0.20	0.02	-0.35
Ca	0.22	0.03	-0.52	0.19	-0.21	-0.15	-0.04	0.33	0.15	0.15	-0.25	0.07	0.18	0.16	0.23	-0.50
Mn	0.35	0.06	0.12	0.09	-0.27	-0.16	0.40	-0.11	-0.41	-0.42	0.01	0.03	0.29	-0.37	0.02	-0.13
Fe	0.21	0.03	0.24	0.43	0.22	0.03	-0.19	-0.44	0.31	-0.24	-0.38	0.28	0.04	0.14	-0.12	-0.09
Co	-0.11	0.24	0.18	0.54	-0.03	0.23	-0.21	-0.01	-0.23	0.50	0.26	0.05	0.17	-0.29	0.07	-0.11
Ni	0.03	0.50	-0.05	0.09	0.02	-0.46	0.01	-0.17	-0.20	-0.04	0.19	0.13	-0.55	0.22	0.23	-0.02
Cu	0.19	0.15	-0.40	-0.02	0.11	0.48	0.16	-0.26	-0.33	0.14	-0.35	-0.22	-0.31	-0.05	-0.22	-0.01

Zn	0.28	-0.11	-0.13	0.52	-0.03	0.00	-0.13	0.39	0.02	-0.27	0.14	-0.34	-0.16	0.07	-0.12	0.44
Se	0.04	-0.53	0.06	0.07	0.05	0.24	-0.06	0.07	-0.49	-0.07	0.08	0.41	-0.14	0.34	0.27	-0.05
Rb	0.19	0.46	0.19	-0.16	-0.12	0.14	0.08	0.32	-0.11	0.14	-0.22	0.33	0.27	0.35	-0.11	0.39
Sr	0.17	-0.08	-0.51	-0.09	-0.21	-0.07	-0.25	-0.39	0.04	0.07	0.34	0.34	0.21	-0.06	-0.18	0.32
Mo	0.02	0.15	-0.15	-0.15	0.64	-0.15	-0.38	0.28	-0.27	-0.21	0.00	0.13	0.15	-0.23	-0.20	-0.13
Cd	-0.07	-0.01	-0.21	0.22	0.53	-0.02	0.65	-0.05	0.14	0.10	0.24	0.07	0.24	0.17	0.08	0.10

IN09	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.83	1.35	1.35	1.22	1.08	1.02	0.93	0.89	0.83	0.77	0.73	0.68	0.65	0.62	0.57	0.49
Proportion of Variance	0.21	0.11	0.11	0.09	0.07	0.07	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.01
Cumulative Proportion	0.21	0.32	0.44	0.53	0.60	0.67	0.72	0.77	0.82	0.85	0.89	0.91	0.94	0.96	0.99	1.00
Mg	-0.27	0.27	-0.21	0.25	-0.04	0.02	-0.49	0.00	-0.16	0.10	0.50	0.17	0.29	0.10	0.31	0.07
P	-0.28	0.09	-0.34	0.19	-0.09	0.37	-0.07	0.18	-0.01	0.20	-0.32	-0.57	-0.12	-0.30	0.09	-0.05
S	-0.34	0.04	-0.08	0.19	0.17	-0.08	0.32	-0.08	-0.48	-0.57	0.04	-0.03	0.13	-0.22	-0.20	0.16
K	-0.34	0.05	0.11	0.02	0.36	0.30	0.35	0.19	-0.13	0.20	0.16	0.13	-0.32	0.44	0.07	-0.29
Ca	-0.02	0.49	0.44	-0.03	-0.13	-0.05	0.07	0.07	-0.13	0.19	-0.05	-0.17	-0.15	0.11	-0.03	0.64
Mn	-0.31	0.11	-0.08	0.08	0.07	-0.52	-0.39	-0.07	-0.09	0.05	-0.46	0.09	-0.24	0.25	-0.25	-0.16
Fe	-0.36	-0.01	-0.08	-0.16	-0.24	-0.20	0.22	-0.10	0.46	-0.11	0.24	-0.42	0.26	0.36	-0.15	-0.03
Co	-0.06	-0.06	-0.02	-0.62	0.24	-0.30	-0.09	0.08	-0.31	0.24	0.32	-0.28	-0.08	-0.30	-0.02	-0.10
Ni	-0.16	-0.34	0.20	0.05	-0.13	-0.06	-0.18	0.83	0.08	-0.15	0.06	0.08	0.03	-0.04	-0.14	0.11
Cu	-0.36	-0.02	0.19	-0.22	-0.14	0.32	-0.02	-0.16	-0.03	0.33	-0.16	0.34	0.41	-0.20	-0.42	-0.05
Zn	-0.40	0.01	-0.08	-0.16	-0.23	-0.18	0.21	-0.07	0.29	-0.04	-0.03	0.38	-0.35	-0.36	0.44	0.12
Se	0.13	0.37	-0.29	-0.35	-0.04	-0.08	0.22	0.36	-0.13	-0.07	-0.36	0.14	0.41	0.18	0.27	-0.09
Rb	-0.13	-0.01	0.31	0.10	0.68	-0.10	-0.06	-0.02	0.35	0.01	-0.20	-0.10	0.34	-0.15	0.28	0.08
Sr	0.03	0.43	0.48	0.12	-0.23	-0.06	-0.03	0.05	0.03	-0.20	0.08	-0.10	-0.01	-0.25	0.03	-0.62
Mo	0.02	0.37	-0.18	-0.34	0.26	0.35	-0.30	0.04	0.33	-0.41	0.08	0.09	-0.24	-0.03	-0.28	0.09
Cd	-0.16	-0.28	0.30	-0.33	-0.16	0.28	-0.30	-0.20	-0.24	-0.35	-0.21	-0.15	-0.01	0.26	0.38	0.01

IN10	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.81	1.46	1.38	1.22	1.15	1.00	0.92	0.84	0.82	0.80	0.70	0.64	0.58	0.55	0.49	0.45
Proportion of Variance	0.20	0.13	0.12	0.09	0.08	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.01
Cumulative Proportion	0.20	0.34	0.46	0.55	0.63	0.70	0.75	0.79	0.84	0.88	0.91	0.93	0.95	0.97	0.99	1.00
Mg	-0.29	-0.04	0.15	-0.40	0.09	-0.23	0.50	-0.07	0.20	-0.20	0.23	-0.06	-0.13	0.49	-0.02	0.15
P	-0.35	0.30	0.15	-0.18	0.03	0.01	0.09	-0.15	-0.10	-0.44	-0.22	-0.19	0.50	-0.35	0.11	-0.20
S	-0.36	0.07	0.17	-0.15	0.22	0.28	-0.06	0.38	0.33	0.25	0.10	-0.30	-0.33	-0.33	-0.17	-0.14
K	-0.27	0.18	0.31	0.22	0.23	0.22	-0.43	-0.07	-0.09	0.19	0.13	-0.13	0.29	0.48	0.13	0.22
Ca	-0.10	-0.40	0.42	0.18	-0.25	-0.03	0.07	0.02	-0.02	-0.08	-0.29	-0.12	0.04	-0.17	-0.36	0.53
Mn	-0.27	-0.32	0.04	-0.06	0.31	-0.37	0.01	-0.18	-0.06	0.39	0.27	0.37	0.24	-0.35	0.07	0.04
Fe	-0.37	-0.01	-0.24	-0.17	-0.30	-0.24	-0.21	0.11	-0.11	0.12	-0.29	-0.13	-0.27	-0.02	0.56	0.26

Co	-0.22	-0.07	-0.31	0.38	-0.23	-0.27	-0.15	-0.27	0.15	-0.16	0.45	-0.43	-0.03	-0.06	-0.20	-0.09
Ni	-0.10	-0.29	-0.29	0.18	0.28	0.31	0.13	-0.39	0.50	0.01	-0.40	-0.01	0.04	0.06	0.16	0.01
Cu	-0.37	-0.06	0.02	0.26	0.13	0.25	-0.04	-0.08	-0.34	-0.43	0.08	0.43	-0.46	-0.04	-0.01	-0.05
Zn	-0.33	-0.06	-0.36	-0.22	-0.18	0.00	-0.26	0.22	0.06	0.02	-0.22	0.30	0.25	0.25	-0.52	-0.15
Se	-0.05	0.46	0.07	0.28	-0.22	-0.10	0.05	0.16	0.55	-0.08	0.10	0.45	0.07	-0.12	0.11	0.26
Rb	-0.14	0.31	0.08	0.41	0.16	-0.41	0.28	-0.01	-0.15	0.29	-0.42	-0.06	-0.16	0.14	-0.18	-0.25
Sr	-0.07	-0.42	0.37	0.19	-0.31	-0.04	0.00	0.24	0.17	-0.04	0.01	0.06	0.11	0.17	0.32	-0.56
Mo	-0.08	0.18	0.23	-0.21	-0.49	0.24	0.02	-0.60	0.00	0.35	0.02	0.12	-0.17	-0.02	-0.10	-0.16
Cd	-0.19	-0.01	-0.28	0.21	-0.22	0.40	0.56	0.23	-0.27	0.26	0.19	-0.05	0.26	0.00	0.09	0.10

NC06	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.94	1.32	1.18	1.11	1.03	1.02	0.98	0.94	0.89	0.83	0.77	0.74	0.66	0.57	0.54	0.47
Proportion of Variance	0.24	0.11	0.09	0.08	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01
Cumulative Proportion	0.24	0.35	0.43	0.51	0.58	0.64	0.70	0.76	0.81	0.85	0.89	0.92	0.95	0.97	0.99	1.00
Mg	-0.34	0.19	-0.06	0.12	-0.51	0.14	-0.04	0.08	-0.11	0.05	-0.06	0.19	0.06	-0.23	-0.60	-0.26
P	-0.37	0.15	-0.09	0.30	-0.22	-0.03	0.12	0.09	-0.29	0.25	0.00	0.17	-0.15	-0.12	0.45	0.51
S	-0.32	-0.04	-0.22	-0.12	0.14	-0.27	-0.22	-0.19	-0.13	0.31	0.11	-0.65	-0.22	-0.03	-0.23	0.06
K	-0.38	0.15	-0.22	-0.10	0.28	-0.15	0.03	0.17	-0.17	0.10	-0.02	0.19	0.10	0.22	0.32	-0.63
Ca	-0.06	0.46	0.40	-0.30	-0.05	0.01	-0.07	-0.21	0.07	-0.13	0.40	0.11	-0.50	-0.10	0.11	-0.11
Mn	-0.25	0.01	0.19	0.24	-0.32	-0.41	-0.12	-0.18	0.04	-0.59	-0.06	-0.23	0.22	0.21	0.16	-0.04
Fe	-0.31	-0.19	0.07	0.09	-0.03	0.41	0.21	-0.15	0.43	0.06	0.10	-0.29	0.17	-0.41	0.29	-0.20
Co	-0.17	-0.25	0.10	-0.55	-0.23	-0.12	-0.03	0.29	0.18	-0.03	-0.56	0.01	-0.28	-0.10	0.12	0.05
Ni	-0.12	-0.27	0.37	-0.09	-0.02	0.29	-0.24	0.55	-0.34	-0.08	0.36	-0.20	0.12	0.12	0.01	0.03
Cu	-0.23	-0.02	0.24	0.13	0.43	0.26	0.36	-0.16	-0.36	-0.31	-0.38	-0.07	-0.24	-0.01	-0.20	0.03
Zn	-0.34	-0.20	0.25	0.05	0.04	-0.03	0.21	-0.06	0.41	0.27	0.13	0.21	-0.05	0.59	-0.23	0.16
Se	-0.10	-0.19	-0.23	-0.55	-0.19	0.19	0.16	-0.47	-0.32	-0.11	0.17	0.13	0.28	0.13	0.04	0.15
Rb	-0.32	0.08	-0.15	-0.15	0.43	-0.17	-0.15	0.18	0.23	-0.32	0.17	0.26	0.24	-0.34	-0.19	0.34
Sr	-0.03	0.50	0.37	-0.17	0.09	0.11	-0.20	-0.10	0.00	0.31	-0.35	-0.14	0.50	0.07	0.03	0.15
Mo	-0.06	0.36	-0.46	-0.01	-0.06	0.48	-0.10	0.17	0.23	-0.28	-0.07	-0.22	-0.16	0.39	0.03	0.13
Cd	-0.11	-0.28	0.00	0.18	0.10	0.25	-0.74	-0.34	0.00	0.03	-0.17	0.30	-0.15	0.02	0.09	-0.05

NY05	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.99	1.44	1.23	1.16	1.04	1.00	0.92	0.88	0.83	0.81	0.75	0.67	0.63	0.56	0.49	0.32
Proportion of Variance	0.25	0.13	0.09	0.08	0.07	0.06	0.05	0.05	0.04	0.04	0.04	0.03	0.02	0.02	0.02	0.01
Cumulative Proportion	0.25	0.38	0.47	0.56	0.62	0.69	0.74	0.79	0.83	0.87	0.91	0.93	0.96	0.98	0.99	1.00
Mg	0.36	-0.05	0.29	-0.15	0.14	0.14	-0.21	0.18	-0.02	0.07	0.27	-0.26	-0.41	-0.09	-0.56	0.10
P	0.39	0.05	0.29	0.12	0.11	-0.05	-0.22	0.09	0.00	0.18	0.05	-0.03	-0.28	0.24	0.71	0.01
S	0.34	0.10	0.10	-0.08	-0.24	0.03	0.08	-0.33	-0.18	0.25	-0.55	0.46	-0.21	0.00	-0.20	-0.03
K	0.19	-0.02	-0.01	0.51	0.42	-0.31	-0.10	-0.21	0.06	0.15	-0.34	-0.37	0.25	-0.05	-0.18	-0.02

Ca	0.12	-0.59	-0.27	-0.14	0.03	0.04	-0.05	-0.04	-0.05	0.00	-0.14	-0.03	0.01	-0.13	0.14	0.69
Mn	0.35	-0.10	0.08	-0.32	0.09	-0.06	-0.11	0.03	-0.17	0.26	0.28	0.21	0.70	0.09	-0.05	-0.11
Fe	0.36	0.10	-0.10	-0.11	-0.23	0.04	0.17	-0.18	-0.01	-0.51	-0.07	-0.34	0.11	0.57	-0.06	0.06
Co	0.17	0.31	-0.27	-0.09	0.21	-0.06	0.15	0.55	-0.53	-0.17	-0.24	-0.06	-0.01	-0.21	0.07	0.00
Ni	0.05	0.21	-0.40	-0.14	0.39	0.46	-0.24	0.12	0.47	0.07	-0.20	0.17	0.01	0.22	-0.05	-0.01
Cu	0.21	-0.05	-0.38	0.47	0.06	-0.22	-0.07	-0.02	-0.09	-0.21	0.41	0.50	-0.18	0.13	-0.15	0.03
Zn	0.36	0.17	-0.15	-0.18	-0.02	-0.05	-0.05	-0.34	0.27	-0.30	0.13	-0.05	0.00	-0.65	0.20	-0.15
Se	0.25	-0.02	-0.03	0.10	-0.37	-0.28	0.31	0.51	0.55	0.18	-0.07	0.02	0.08	-0.05	-0.06	0.07
Rb	0.15	-0.12	0.10	0.19	0.29	0.46	0.75	-0.10	-0.01	0.09	0.17	0.02	0.00	-0.07	0.07	-0.04
Sr	0.08	-0.62	-0.24	-0.09	-0.03	0.01	-0.04	0.10	-0.05	-0.01	-0.14	-0.11	-0.17	0.04	0.02	-0.68
Mo	0.09	-0.14	0.31	0.42	-0.24	0.48	-0.30	0.23	-0.03	-0.35	-0.17	0.13	0.28	-0.18	-0.01	-0.03
Cd	0.07	0.18	-0.43	0.23	-0.44	0.31	-0.11	-0.09	-0.20	0.47	0.19	-0.34	0.02	-0.06	0.03	0.00

NY06	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	2.05	1.57	1.22	1.15	1.14	0.96	0.94	0.83	0.78	0.68	0.68	0.62	0.56	0.50	0.44	0.27
Proportion of Variance	0.26	0.15	0.09	0.08	0.08	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.01	0.00
Cumulative Proportion	0.26	0.42	0.51	0.59	0.67	0.73	0.79	0.83	0.87	0.90	0.92	0.95	0.97	0.98	1.00	1.00
Mg	-0.37	0.16	-0.08	0.06	-0.33	0.11	-0.24	0.08	-0.03	0.12	0.11	-0.37	0.00	0.51	0.25	-0.39
P	-0.40	0.05	-0.03	0.24	-0.17	-0.14	0.28	-0.14	-0.12	0.03	-0.08	-0.16	-0.34	-0.14	0.38	0.55
S	-0.42	0.13	-0.17	-0.02	0.05	-0.04	0.00	-0.18	-0.05	-0.05	-0.07	-0.16	-0.37	-0.34	-0.60	-0.31
K	-0.30	-0.13	0.13	0.23	0.01	-0.45	0.19	0.29	0.31	-0.07	-0.51	0.16	0.24	0.07	0.00	-0.22
Ca	-0.11	-0.16	-0.27	-0.64	0.09	0.17	-0.07	-0.15	0.23	0.01	-0.47	-0.24	0.11	0.10	0.05	0.23
Mn	-0.29	0.06	-0.11	-0.24	-0.26	0.27	0.42	0.28	0.08	-0.39	0.35	0.08	0.34	-0.21	0.00	0.00
Fe	-0.34	-0.08	0.15	0.06	0.23	0.13	-0.41	-0.38	0.24	-0.09	0.13	0.24	0.15	-0.36	0.40	-0.17
Co	0.03	-0.05	0.58	-0.29	-0.35	-0.17	-0.13	-0.17	0.10	-0.48	0.01	0.05	-0.28	0.18	-0.10	0.05
Ni	-0.18	0.27	0.36	-0.36	-0.09	0.04	-0.14	0.43	-0.04	0.55	-0.04	0.23	-0.12	-0.22	0.01	0.07
Cu	-0.20	-0.28	0.30	-0.10	0.06	0.16	0.46	-0.42	-0.35	0.31	-0.02	0.16	0.16	0.23	-0.07	-0.17
Zn	-0.31	-0.32	0.07	0.21	0.15	0.14	-0.21	0.13	0.27	0.11	0.30	0.02	0.06	0.30	-0.42	0.44
Se	-0.04	-0.33	-0.14	0.11	-0.51	0.03	-0.39	0.03	-0.48	-0.04	-0.23	0.15	0.26	-0.21	-0.13	0.10
Rb	-0.15	0.44	-0.34	-0.06	0.05	-0.07	-0.07	-0.15	-0.12	-0.13	-0.06	0.65	-0.04	0.38	-0.04	0.13
Sr	0.01	-0.52	-0.20	-0.04	0.04	0.24	0.06	0.30	0.01	-0.07	-0.03	0.30	-0.59	0.06	0.20	-0.22
Mo	0.21	0.05	-0.11	0.17	-0.54	0.22	0.14	-0.30	0.56	0.31	-0.07	0.18	-0.05	-0.06	-0.10	-0.03
Cd	-0.01	-0.27	-0.28	-0.31	-0.13	-0.68	-0.01	-0.11	0.06	0.23	0.44	0.05	0.02	-0.03	0.05	-0.05

NY12	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	1.89	1.42	1.28	1.18	1.05	0.98	0.96	0.94	0.87	0.81	0.78	0.67	0.61	0.53	0.53	0.31
Proportion of Variance	0.22	0.13	0.10	0.09	0.07	0.06	0.06	0.06	0.05	0.04	0.04	0.03	0.02	0.02	0.02	0.01
Cumulative Proportion	0.22	0.35	0.45	0.54	0.61	0.67	0.73	0.78	0.83	0.87	0.91	0.94	0.96	0.98	0.99	1.00
Mg	0.34	-0.16	0.03	-0.21	0.17	-0.29	0.42	-0.16	0.12	0.16	-0.19	0.30	-0.02	-0.47	0.34	0.01

P	0.39	0.02	0.19	0.15	0.06	0.05	0.17	-0.33	0.21	-0.08	-0.16	0.37	0.03	0.53	-0.38	0.03
S	0.31	-0.09	-0.25	-0.20	0.01	0.22	-0.34	0.07	-0.10	-0.44	-0.41	0.03	-0.50	-0.08	0.02	0.02
K	0.14	0.05	0.44	0.35	0.39	0.16	-0.20	0.07	-0.19	-0.37	-0.02	0.05	0.36	-0.10	0.36	0.00
Ca	0.16	0.60	0.05	-0.24	0.07	-0.05	-0.08	0.05	-0.11	0.16	0.03	-0.02	-0.03	0.05	0.05	0.70
Mn	0.37	-0.07	0.10	-0.26	0.10	-0.05	0.24	-0.11	0.04	-0.13	0.04	-0.80	0.10	0.16	0.01	-0.09
Fe	0.37	-0.16	-0.17	-0.04	-0.31	0.02	-0.06	0.24	-0.18	0.15	0.31	0.18	0.04	0.44	0.51	-0.08
Co	-0.01	0.23	-0.16	0.32	-0.07	0.18	0.60	0.56	0.00	-0.08	-0.32	-0.06	-0.02	0.05	0.01	0.00
Ni	-0.02	0.07	-0.41	0.03	0.59	0.08	0.01	0.08	0.49	-0.12	0.41	0.05	-0.12	0.08	0.09	0.00
Cu	0.24	0.02	0.15	0.54	0.15	-0.26	-0.11	0.02	-0.14	0.35	0.11	-0.18	-0.56	-0.06	-0.06	-0.08
Zn	0.36	-0.19	-0.16	-0.05	0.12	0.30	0.06	0.18	-0.36	0.12	0.32	0.10	0.22	-0.32	-0.51	0.06
Se	0.19	0.14	-0.08	0.28	-0.37	0.54	-0.08	-0.32	0.39	0.17	0.02	-0.15	0.07	-0.29	0.16	0.05
Rb	0.04	-0.23	0.37	-0.23	0.15	0.20	-0.25	0.47	0.39	0.44	-0.25	-0.02	0.02	0.07	-0.05	-0.02
Sr	0.11	0.63	0.02	-0.24	0.07	0.05	-0.09	0.00	-0.09	0.13	-0.02	0.11	0.03	-0.05	-0.03	-0.69
Mo	0.25	0.11	0.14	0.05	-0.38	-0.43	-0.16	0.33	0.39	-0.37	0.22	0.05	0.08	-0.23	-0.19	-0.01
Cd	0.14	-0.01	-0.51	0.23	0.11	-0.34	-0.31	-0.01	0.02	0.19	-0.41	-0.12	0.47	0.04	-0.04	0.01

MO06	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	2.01	1.48	1.30	1.20	1.11	1.04	0.89	0.78	0.77	0.74	0.70	0.67	0.58	0.49	0.37	0.33
Proportion of Variance	0.25	0.14	0.10	0.09	0.08	0.07	0.05	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.01	0.01
Cumulative Proportion	0.25	0.39	0.50	0.59	0.66	0.73	0.78	0.82	0.86	0.89	0.92	0.95	0.97	0.98	0.99	1.00
Mg	-0.26	0.12	-0.02	0.05	-0.25	0.54	-0.28	0.08	-0.16	0.09	-0.60	0.12	-0.24	-0.01	0.00	-0.08
P	-0.41	0.06	-0.09	0.04	0.01	0.24	0.33	0.05	0.01	-0.15	0.03	0.05	0.37	0.35	-0.43	0.42
S	-0.30	0.26	-0.05	0.16	0.25	-0.29	0.16	-0.13	-0.15	-0.48	-0.08	-0.01	-0.53	0.22	0.21	0.03
K	-0.38	0.03	0.15	0.12	-0.05	0.11	0.52	-0.09	0.15	-0.04	-0.05	-0.03	0.25	-0.47	0.28	-0.37
Ca	-0.16	-0.51	0.15	0.07	0.34	0.06	0.00	0.08	-0.20	0.01	0.12	0.21	-0.18	0.08	-0.42	-0.48
Mn	-0.22	0.10	-0.24	0.45	0.04	0.29	-0.29	-0.01	-0.24	0.16	0.56	-0.20	0.06	0.04	0.25	-0.05
Fe	-0.33	0.11	0.13	-0.22	-0.02	-0.35	0.04	-0.04	-0.18	0.61	-0.03	0.26	0.11	0.37	0.25	-0.07
Co	-0.11	-0.27	-0.44	0.13	-0.12	-0.31	0.22	0.49	-0.20	0.23	-0.18	-0.29	-0.15	-0.22	-0.06	0.15
Ni	-0.08	-0.29	-0.52	-0.25	0.17	-0.04	-0.23	-0.13	0.00	-0.29	-0.24	-0.06	0.42	0.18	0.26	-0.23
Cu	-0.22	-0.07	-0.01	-0.52	-0.25	0.17	0.00	0.45	0.28	-0.18	0.38	0.02	-0.26	0.11	0.18	-0.08
Zn	-0.24	0.28	-0.35	-0.32	0.14	-0.11	-0.20	-0.19	-0.04	0.05	0.16	0.32	-0.07	-0.54	-0.30	0.09
Se	-0.21	0.17	0.18	0.35	-0.01	-0.36	-0.43	0.45	0.32	-0.18	-0.08	0.17	0.26	-0.04	-0.09	-0.10
Rb	-0.30	-0.22	-0.06	0.10	-0.26	-0.15	-0.15	-0.47	0.52	0.17	-0.01	-0.33	-0.22	0.10	-0.21	-0.03
Sr	-0.19	-0.52	0.24	0.06	0.17	0.05	-0.16	-0.07	0.04	-0.03	-0.02	0.24	-0.04	-0.22	0.35	0.58
Mo	-0.22	0.02	0.43	-0.31	-0.06	-0.12	-0.23	-0.03	-0.44	-0.16	0.00	-0.56	0.15	-0.15	-0.13	0.04
Cd	-0.04	0.19	0.04	-0.14	0.72	0.20	-0.02	0.19	0.33	0.29	-0.15	-0.35	-0.06	0.00	0.03	0.05

SA10	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16
Standard deviation	2.37	1.57	1.19	1.07	1.01	0.93	0.83	0.75	0.71	0.69	0.60	0.54	0.46	0.42	0.35	0.28

Proportion of Variance	0.35	0.15	0.09	0.07	0.06	0.05	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01	0.00
Cumulative Proportion	0.35	0.51	0.59	0.67	0.73	0.78	0.83	0.86	0.89	0.92	0.95	0.96	0.98	0.99	1.00	1.00
Mg	0.31	-0.12	0.24	-0.17	0.39	-0.07	0.12	-0.32	0.20	-0.09	0.16	-0.16	0.14	-0.38	-0.34	-0.39
P	0.32	-0.21	0.31	0.12	0.07	0.11	-0.18	-0.27	0.08	0.11	0.03	0.05	0.17	0.00	0.75	0.10
S	0.29	-0.19	0.15	-0.12	0.09	-0.15	-0.27	0.58	0.09	-0.31	-0.22	0.43	0.01	-0.23	-0.06	0.11
K	0.24	-0.12	0.36	0.15	-0.49	0.09	0.11	-0.03	-0.05	-0.31	-0.32	-0.50	-0.04	0.02	-0.18	0.18
Ca	0.13	0.49	0.23	-0.06	0.12	-0.16	0.29	-0.04	-0.19	-0.11	-0.27	0.13	-0.54	-0.01	0.26	-0.26
Mn	0.36	0.15	0.07	-0.09	0.18	0.18	0.06	-0.23	0.13	0.29	-0.02	0.23	-0.20	0.20	-0.34	0.60
Fe	0.30	0.09	-0.08	-0.34	-0.01	-0.31	-0.01	0.25	-0.22	-0.07	0.58	-0.38	-0.10	0.07	0.15	0.21
Co	0.26	0.20	-0.37	-0.12	-0.20	-0.02	-0.22	0.10	0.27	0.47	-0.32	-0.27	-0.07	-0.38	0.09	-0.09
Ni	0.16	0.28	-0.23	0.36	0.17	0.06	-0.65	-0.25	-0.26	-0.34	0.03	-0.07	-0.07	-0.01	-0.12	-0.03
Cu	0.29	-0.13	-0.08	0.10	-0.13	0.33	0.20	0.09	-0.70	0.25	0.07	0.21	0.12	-0.26	-0.09	-0.14
Zn	0.36	-0.19	-0.12	-0.04	0.10	0.02	-0.06	0.15	0.03	0.12	-0.17	-0.07	0.03	0.73	-0.08	-0.45
Se	-0.11	0.29	0.33	0.03	0.17	0.66	-0.10	0.43	0.16	0.05	0.21	-0.22	-0.04	0.00	0.01	-0.10
Rb	0.25	0.19	-0.04	0.21	-0.56	0.02	0.06	-0.06	0.37	-0.11	0.44	0.37	-0.05	0.02	-0.05	-0.23
Sr	0.08	0.56	0.07	-0.10	-0.01	-0.12	0.12	0.03	-0.06	-0.06	-0.16	0.05	0.76	0.11	0.00	0.06
Mo	-0.17	0.05	0.55	0.00	-0.20	-0.35	-0.42	-0.01	-0.18	0.46	0.08	0.06	-0.02	0.06	-0.21	-0.13
Cd	0.10	0.01	0.00	0.76	0.27	-0.33	0.25	0.27	0.10	0.20	0.07	-0.14	0.03	-0.05	-0.03	0.10

Table S4. Weather Station Locations.

Location	**Weather Station**
Florida	Homestead General Aviation Airport
Indiana	West Lafayette 6 NW
North Carolina	Clayton Field
New York	Aurora Research Farm
Missouri	Columbia U of M

CHAPTER 4:
EXPRESSION ANALYSIS IN MAIZE ROOTS DESCRIBES GENE REGULATORY RELATIONSHIPS AND IDENTIFIES CANDIDATE GENES FOR PREVIOUSLY MAPPED LEAF AND SEED IONOME QTL

Alexandra Asaro, Jennifer Barrett, Rob Schaefer, Brian P. Dilkes, Ivan Baxter

This chapter will be submitted for publication in a peer-reviewed journal.
AA conducted all data analysis other than co-expression analysis. AA interpreted results, created figures, and wrote the manuscript with assistance from BD and IB. JB collected ionomics and RNA sequencing data. RS performed co-expression analysis.

ABSTRACT

Roots of young plants undergo highly regulated gene expression changes that pattern root architecture and physiology, with lifelong effects on the structural integrity, water-use efficiency, and nutrient flow of the plant. Many phenotypes, such as seed and leaf element accumulation are often determined by gene expression in the root. To understand gene regulatory networks in maize roots, we measured transcript levels in two-week-old roots of 218 greenhouse-grown plants belonging to the maize Intermated B73 x Mo17 (IBM) recombinant inbred population. We also profiled the ionome of leaf samples from the same plants and carried out QTL mapping on 20 element traits. After performing quality control on the root RNA-seq data, we retained an average of 19.6 million reads per sample. Following quantification with an alignment bias-reducing pipeline, gene expression estimates were used for expression QTL (eQTL) mapping and co-expression analysis which identified 12,497 cis-eQTL, 6,128 trans-eQTL, and 250 co-expressed gene clusters. We detected 8 trans-eQTL hotspots, and found significantly enriched co-expression and gene ontology among hotspot gene targets. Finally, we performed a correlation analysis between root gene expression and leaf element measurements. For 10 elements, genes where root expression correlated with leaf element content co-located with leaf QTL mapped for the element. Additionally, for cadmium and zinc, correlated genes on different chromosomes had trans-eQTL mapping back to the element QTL. The chromosome 2 locus associated with both leaf and seed cadmium content co-localizes with the trans-eQTL hotspot on chromosome 2, which has among its gene targets the top 5 cadmium-correlated genes outside of the QTL interval. Dissecting these relationships can aid in understanding mechanisms and candidate genes underlying element accumulation QTL detected in the leaf and seed.

INTRODUCTION

In this study, we used the maize Intermated B73 x Mo17 (IBM) recombinant inbred population to conduct an analysis of gene expression in maize roots and connect gene expression in the roots with leaf and seed ionome phenotypes. Gene regulation in the roots has a strong impact on element accumulation throughout the plant, making root gene expression an ideal resource for understanding biological changes that cause ionome variation in leaf and seed tissue [1–3]. We can connect variation in root gene expression with variation in the ionome using co-expression and expression QTL (eQTL) analyses. Expression QTL mapping follows the same principles as standard QTL mapping, with the distinguishing characteristic being that the trait of interest is gene expression [4]. Expression QTL for a given transcript located within or near the gene encoding that transcript are referred to as cis-eQTL, while eQTL distantly located from the transcript they regulate are considered trans-eQTL [5]. Co-expression analysis tests whether genes contributing to a trait operate in a co-regulated network. If candidate genes for a trait are co-regulated, they should be more co-expressed than a random set of genes the same size and candidate genes most highly co-expressed are likely to be the causal genes [6]. Consistent with roots being a key regulatory source for the ionome, a recent co-expression study revealed that candidate genes for kernel element SNPs were more co-expressed in root expression networks than in networks derived from other tissues [7].

To achieve accurate estimates of gene expression in a population with genetic diversity, it is necessary to address the issue of alignment bias based on reference genome. RNA-seq reads must be aligned to a reference genome as a first step in gene quantification. The choice of reference genome can cause alignment bias and dramatically influence results of downstream analyses. For example, in eQTL analysis, the genome is surveyed for associations between

parental genotype and expression levels of different transcripts. This analysis may detect a false positive association between the reference allele and increase in expression of a gene, even if gene expression is the same across the population, if the lines with the reference allele just aligned better to the reference genome than those with the non-reference allele. Several previous studies in maize align reads using a single reference genome and do not address alignment bias [8, 9]. In older experiments, some noted a bias but did not have many resources to address the problem. Holloway et al. used microarrays to conduct an eQTL experiment in maize inbreds and observed a substantial bias toward the reference allele, with a larger proportion of cis-eQTL having higher reference allele expression than would be expected without bias. The false positive cis-eQTLs confirmed were often some of the strongest cis-eQTLs mapped [4]. To date, there has been no well-tested and standardized method for dealing with mapping bias in a bi-parental RIL population.

Predicting and accounting for the variety of scenarios that can cause mapping bias in a species as diverse as maize and in a bi-parental population with extensive recombination has numerous complications. We tested several methods to account for mapping bias and, while each method to reduce mapping bias is imperfect, employing the reference of B73 along with consideration of Mo17 polymorphisms was determined to be the most functional and reasonably executable approach with the least drawbacks.

Here, we have estimated expression of genes expressed in two-week-old maize roots, modeled relationships between genetic variation and gene expression, and determined co-expressed gene modules. We related genotype to phenotype through several levels of analysis, connecting element accumulation with gene expression in the root. This integrative approach has

allowed us to identify candidate genes for previously mapped ionome QTL and genes that are functionally connected with these QTL.

MATERIALS & METHODS

Population and Data Collection

Greenhouse growth and sampling. 227 Intermated B73 x Mo17 (IBM) recombinant inbred population lines as well as the B73 and Mo17 parent lines were grown in a greenhouse with the following growth conditions: Day temp: 26-28C, Night temp: 22-24C, 14-hour day, 50% relative humidity. Leaf and root sampling was performed two weeks after planting. The youngest, fully expanded leaf was taken, dried down, then crumbled into a tube for ionomics. The roots were cut off at the stem, and then a 1-inch segment of the root was removed for RNA-sequencing, 1 inch below the base of the stem. Samples were immediately placed in liquid nitrogen and then ground using mortars and pestles.

Elemental profile analysis. Elemental profile analysis was conducted on leaf samples following a standardized pipeline in the Baxter Lab with the same methods as reported in Veley et al. [10]. Descriptions taken directly are marked in quotations. Samples were "weighed into borosilicate glass test tubes and digested in 2.5 ml nitric acid (AR select, Macron) containing 20 ppb indium as a sample preparation internal standard. Digestion was carried out by soaking overnight at room temperature and then heating to 95°C for 4 hrs. After cooling, samples were diluted to 10 ml using ultra-pure water (UPW, Millipore Milli-Q). Samples were diluted an additional 5 × with UPW containing yttrium as an instrument internal standard using an ESI prepFAST autodilution system (Elemental Scientific). A PerkinElmer NexION 350D with helium mode enabled for improved removal of spectral interferences was used to measure

concentrations of," B, Na, Mg, Al, P, S, K, Ca, Fe, Mn, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo, and Cd. "Instrument reported concentrations are corrected for the yttrium and indium internal standards and a matrix-matched control (pooled leaf digestate) as described [11]. The control was run every 10 samples to correct for element-specific instrument drift. Concentrations were converted to parts per million (mg analyte/kg sample) by dividing instrument reported concentrations by the sample weight."

RNA extraction and sequencing. RNA was extracted from root samples using Trizol reagent. RNA from two plants per line was pooled to make a single sample. After removal of low quality preparations, 218 RIL samples (1 sample of each) and samples from the B73 and Mo17 parents (3 of each) were sent for library preparation. Libraries were prepared by Global Biologics using TruSeq RNA Directional (RNAseq) protocol, with a Tru-Seq adapter and a Tru-Seq index ligated to each sample. After library preparation, 8 samples were pooled together (by concentration), to make 28 lanes. 28 pools containing single-end reads were sequenced on one Illumina HiSeq 2500 V4 lane for 100nt, producing an average of over 200 million single-reads per lane. The following procedures were used: Qubit, NO DNA chip (average size: 300bp), dilute to 5nM if above 10nM, qPCR, reads are not paired-end (1 read per sample). The total result of RNA-seq was 5.7 billion 100 base pair-long reads across 224 samples, an average of 25.6 million reads per sample.

Leaf QTL Mapping

Initial element data was prepared for QTL mapping. This data included 421 samples, phenotyped for sample weight and 20 elements, from 227 IBM lines, with replicates ranging from 1-3 replicates (35 lines with 1 replicate, 190 lines with two replicates, 2 lines with 3 replicates). Element phenotypes were weight normalized. 14 samples with a sample weight

below 20 mg were removed. Outliers were removed using the same technique as was used in Asaro et al. for seed element QTL mapping [12], with measurements excluded if the mean absolute deviation (MAD) exceeded 6.2 [13]. Heritability was calculated after outlier removal but before averaging for seed weight and each element using lines with 2 or more replicates. The lmfit and anova functions were implemented to obtain the variances for the genetic component and the residuals. Broad-sense heritability was calculated as the proportion of total variance (genetic plus residuals) explained by the genetic component. After outlier removal, replicates were averaged leaving 225 unique IBM lines. The same IBM genotypes as used in Asaro et al. were merged with phenotypes. Stepwise QTL mapping on 21 traits (20 elements and weight) was conducted in R/QTL [14] using the same method as used in Asaro et al. for seed element mapping. Significance was determined using 1000 random permutations and a 95th percentile LOD score significance threshold.

RNA-Sequencing Data Quality Control, Alignment, and Gene Quantification.

Initial data processing and quality control. RNA-sequencing reads from 224 samples (1 sample from 218 RILs and 3 samples from each parent) were initially processed for quality control using the programs FastQC [15] and Trimmomatic [16]. FastQC was first executed on untrimmed raw FASTQ files, after which quality control results were summarized and assessed. Trimmomatic was then used to trim adapter sequences, trim low quality sequence, and remove low quality reads. Trimmomatic parameters were set to use a 4-mer sliding window, threshold quality score of 15, removal of first 13 bases, removal of adapter sequences, and a minimum length of 36 base pairs (SLIDINGWINDOW:4:15, HEADCROP:13, ILLUMINACLIP:2:30:10, LEADING:3, TRAILING:3, MINLEN:36). Trimming and filtering resulted in an average of 24.9 million reads per sample, with read length ranging 36 to 87 base pairs. FastQC was repeated

on trimmed reads. Ribosomal RNA sequences obtained from SILVA [17] were used to remove any rRNA contamination. Bowtie2 [18] was used to map reads to rRNA sequences and only unmapped reads were carried on for further analysis. Three samples were removed from further analysis because of low total read count, leaving a total of 221 samples. A total of 4.4 billion reads were retained after trimming, filtering, removing rRNA sequences, and removing the samples with low read count, with an average of 19.9 million reads per sample.

Sample validation. Sample identity was tested by calling SNPs using full-genome alignments to the B73 reference genome version 3 [19], aligned with Tophat2 [20] (default parameters), and the program VarScan [21]. SNPs called on the RNA-seq data were compared with SNPs used in previous work with this population [12] to confirm sample identity. To convert previous SNPs from centi-Morgan positions to base pairs, records from Ganal et al. [22] were used to match SNPs to SS numbers, which were entered into dbSNP [23] to look up RS numbers. The batch query service on dbSNP was used with RS numbers to obtain base pair coordinates corresponding to the B73 version 3 reference for each SNP. VarScan was run on RNA-seq alignments with the parameters --min-coverage 20 --min-var-freq 0.08 --p-value 0.05 --output-vcf 1. VCF output files were filtered using VCFtools [24] with parameters --maf 0.1 --max-missing 0.7 --recode --recode-INFO-all (except for chromosome 2 which required less stringent filtering with --maf 0.05 --max-missing 0.4 for adequate coverage). New SNP calls were compared to previous SNPs, with heterozygote calls masked as missing. SNP calls were imputed between the two closest VarScan calls if a VarScan call was not present at the reference SNP location. A distance matrix was built using new and old SNP calls to cluster samples and generate a visual representation. Samples were considered validated as the correct RIL or parent if new SNPs matched previous SNPs with an accuracy above 90% once low accuracy SNPs

(<85% accuracy) were removed. Samples that could not be validated or re-assigned with the correct name were renamed with arbitrary names different from the original IBM line names.

Alignment. The programs Tophat2 and WASP [25] were used to align RNA-sequencing reads to the B73 version 4 reference transcriptome [26]. Parent samples were also aligned to a Mo17 reference transcriptome [27] in Tophat2, again using transcriptome only and two mismatches, to consider the extent of alignment bias. WASP was added to the B73 v4 mapping pipeline to reduce alignment bias. Reads were first mapped to the B73 v4 reference transcriptome with Tophat2. Parameters were set to specify transcriptome-only mapping and a default of two mismatches. Sorted BAM output files were used in the WASP script "find_intersecting_snps.py" along with a set of Mo17 polymorphisms, 8.04 million SNP and insertion/deletion (indel) variants with 164 thousand CDS variants, developed by Peng Zhou and available on the Data Repository for the University of Minnesota [28]. The "find_intersecting_snps.py" script detected reads that intersected with Mo17 SNPs and filtered out reads that overlapped indels, producing a "reads to remap" BAM file, a "reads to keep" BAM file, and a "reads to remap" FASTQ file with reads intersecting SNPs edited to contain the Mo17 allele. The reads that intersected SNPs were then aligned again using Tophat2 with the same parameters as the first alignment pass. The WASP script "filter_remapped_reads.py" was used to remove reads that map to a different location in the genome when SNPs are switched to the Mo17 allele. Reads that mapped to the same location were merged with the "reads to keep" file to generate a sorted, indexed BAM file ready for quantification. This process was repeated for each RIL and parent sample. Parent sample alignment rates were compared with and without inclusion of WASP to assess the reduction in bias achieved by adding WASP to the pipeline.

Quantification. Output files from alignment with Tophat2 and WASP were used in Cufflinks2 to assemble and quantify transcripts. Cufflinks2 was used with the –G option for assembly based on the reference annotation and quantification only of known genes. Genes on chromosomes 1-10 were retained for analysis (genes annotated to contigs were removed). All other parameters were set as default.

Gene Expression Analyses

Leaf ionome and root gene expression correlation tests. Gene expression measurements from 146 RILs that were validated to have the correct IBM line name were used for correlation analysis with leaf ionome data. Genes expressed in less than 80% lines and genes with an expression mean of less than 0.5 RPKM were removed, resulting in 26386 genes for analysis. The Pearson correlation coefficient was calculated between all pairs of genes and leaf elements.

Expression QTL mapping. Gene expression measurements from 215 RILs were used for eQTL mapping. Genes expressed in less than 80% lines and genes with an expression mean of less than 0.5 RPKM were removed, resulting in 26,440 genes for analysis. The SNP set used in Asaro et al. for seed element QTL mapping was converted to from B73 version 3 to version 4 coordinates using the dbSNP archive. The SNP calls made using VarScan (also converted to v4 coordinates) at the positions of previously used SNPs were used as genotypes, allowing for inclusion of lines not previously genotyped. SNPs were filtered to removed SNPs with over 20% missing data and SNPs with below 30% minor allele frequency or above 70% minor allele frequency, resulting in 3,013 SNPs for analysis. eQTL mapping was carried out in the R package Matrix eQTL [29] with parameters useModel = modelLINEAR, pvOutputThreshold_cis = 2e-10, pvOutputThreshold_tra = 1e-10, and cisDist = 1e6. Gene expression was normalized prior to

mapping using quantile normalization to normally distribute measurements while keeping relative rankings. To reduce redundancy from linked SNPs, eQTL were pruned into SNP windows using hierarchical clustering [30] described as follows. First cis and trans eQTL were merged into a single table. For each unique eGene, pairwise correlations were calculated between all eSNPs for that gene. Hierarchical clustering was performed with the R function hclust (method = "complete") and the cutree function was used to define clusters with distance cutoff set at h= 0.4375 (1 – R^2 where R=0.75) to reflect a pairwise correlation cutoff of 0.75. Trans-eQTL hotspots were identified by using the pruned eQTL windows and iterating through every SNP to test for inclusion in an eQTL window. A SNP was given a trans-eQTL count for each instance in which it was present in an eQTL window on a different chromosome or in a window over 15 Mb away on the same chromosome. A SNP was considered a trans-eQTL hotspot if its eQTL count exceeded the 95th percentile of counts across all SNPs in the genome.

Co-expression analysis. Co-expression analysis was carried out on the root gene expression dataset using the python library Camoco (Co-analysis of molecular components) [7]. Expression levels of 38,639 genes across 221 samples (RILs and parent samples) were used as input. Prior to analysis, the following filters were applied: minimum expression level below 0.01 set to NaN, genes missing more than 20% of data removed, accessions missing more than 30% of data removed, and genes must have an expression of 5 RPKM in at least one accession. Gene clusters were calculated using the Markov Cluster (MCL) algorithm on the co-expression matrix. Network health was evaluated and confirmed by testing for normal distribution of raw correlation coefficients and transformed correlation coefficients, balanced clustered gene expression across the genome, and balanced genes removed during QC step across the genome. Enrichment for co-expression among GO term genes was tested including checking to confirm

that no bias occurred based on GO term size. The most recent gene ontology (GO) terms corresponding to the B73 version 4 reference annotation [31] were obtained through MaizeGDB. A 2D visual representation of clusters was generated using ForceAtlas2 [32].

Trans-eQTL hotspot gene ontology and MCL cluster enrichment tests. The target gene sets of trans-eQTL hotspots were tested for enrichment of GO terms and for enrichment of MCL clusters. A hypergeometric calculation followed by correction for multiple testing was used to assess if hotspot gene targets were present in gene lists belonging to MCL clusters or specific GO terms at a level higher than would be randomly expected. The hypergeometric calculation returns p-values for finding a given number of genes in a set of a particular size based on the total number of genes belonging to the type of interest and the total number of genes in the genome.

RESULTS

Population and Growth

For this study, 227 RILs from the IBM population, a population generated through multiple rounds of intermating between the diverse parents Mo17 and B73 followed by several generations of single seed descent, were grown in a greenhouse along with Mo17 and B73. Two weeks after planting, leaves were sampled for ionomics and roots were sampled for root gene expression.

Genetic Control of Leaf Element Concentration

Sampling of 227 IBM lines for leaf ionomics produced 421 samples, each with 20 element measurements. Replicates per line ranged from one to three replicates (35 lines with one replicate, 190 lines with two replicates, 2 lines with three replicates). The youngest, fully

expanded leaf was weighed and profiled for ionomics using inductively-coupled plasma mass spectrometry (ICP-MS). Element measurements were normalized by dividing by sample weight and outliers were removed before averaging line replicates. Heritability was generally high across elements, with broad-sense heritability (H^2) ranging from 0.61 to 0.94 (Table S1). QTL mapping on sample weight and 20 elements was performed using forward/backward regression with the *stepwiseqtl* function in R/QTL [14] and the same genotypes [33] used for our previous seed element mapping study [12]. A significance threshold for QTL was established by setting the stepwise model penalty score as the 95^{th} percentile LOD score achieved across 1000 random *scanone* permutations [34]. 13 total QTL were identified for 12 elements (one QTL for Na, Al, S, K, Ca, Co, Ni, Cu, Zn, As, and Mo and two QTL for Cd) (Table S2). Four loci detected as QTL in the leaf, for molybdenum, cadmium, nickel, and zinc, were collocated with QTL for the same element measured in the seed in field environments [12] (Fig 1). The loci were all loci detected in the seed in multiple environments (3 or more field environments). It is not surprising that these seed element QTL were the ones reproduced with leaf element concentrations from a greenhouse environment, as these loci detected in multiple varying field environments likely have stronger effects and/or lower levels of QTL by environment interactions compared to QTL detected in only one or two environments.

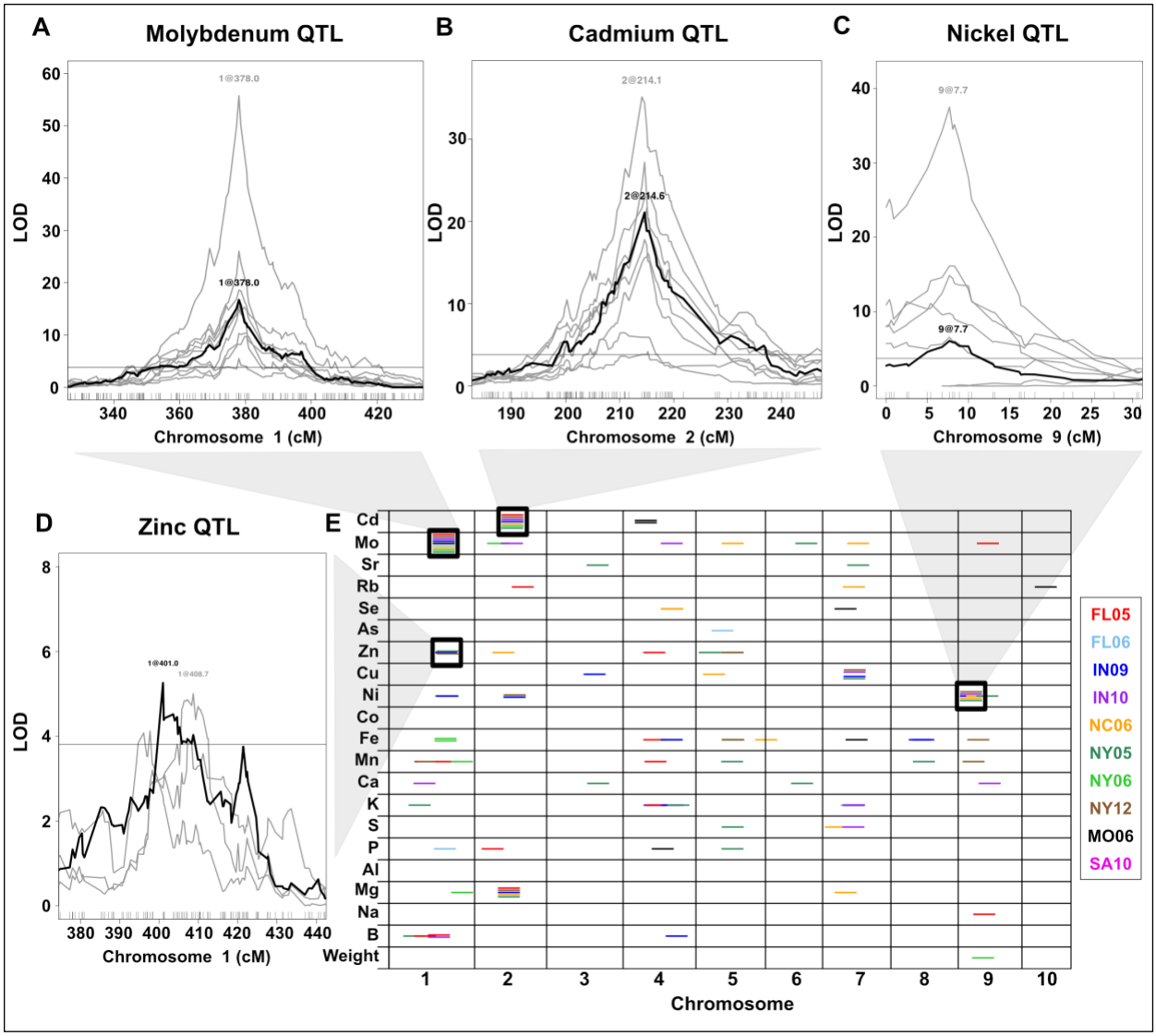


Fig 1. Leaf and Seed Ionome QTL Overlap. (A-D) LOD profile traces at loci with both leaf and seed QTL. Chromosome intervals are shown on the x-axis in centi-Morgans (cM). LOD score is shown on the y-axis and horizontal line is the significance threshold from 1000 random permutations (α= 0.05). Black lines correspond to the QTL mapped in the leaf, grey lines correspond to the QTL mapped in the seed from field environments. (E) Significant previously detected seed element QTL (α= 0.05) from Asaro et al. (2016) shown across the 10 maize chromosomes (in cM). Dashes indicate QTL, colored by environment. All dashes are the same length for visibility. The black boxes around dashes correspond to loci with QTL detected in the leaf as shown in the QTL plots in (A-D).

Root Transcriptome Profiling and Quality Control

RNA was extracted from roots of the IBM RILs and IBM parents, with RNA from two plants per line pooled before RNA sequencing. After removing low quality RNA preparations, one preparation each for 218 RILs and three preparations each of the B73 and Mo17 parent lines were sequenced. Raw RNA-seq data consisted of 5.7 billion 100 base pair-long reads across the 224 samples, with an average of 25.6 million reads per sample. Quality control was applied to ensure high quality reads, remove adapter sequences, and remove reads aligning to ribosomal RNA sequences. Trimming and filtering improved minimum average Phred quality scores from 32 to 36, equivalent to a base call accuracy exceeding 99.9%. After quality control steps and removing three RILs with low read count, a total of 4.4 billion reads were retained across 221 samples, with an average of 19.9 million reads per sample.

In order to verify sample identity, SNPs were called from the RNA-seq data and compared to a SNP set [33] used in prior QTL mapping [12]. SNPs were determined by first aligning reads to the B73 v3 reference genome using Tophat2 [20] and default parameters, and then calling SNPs using VarScan [21]. Filtered newly-called SNPs were compared with the previous SNP set to test for correct sample labeling.

Through this comparison, 146 RILs were validated, 6 RILs were predicted to be switched during labeling, 34 RILs did not match with previous genotypes, and 32 RILs had no previous genotypes for comparison. All three B73 samples were confirmed as B73, however only two of the Mo17 samples were confirmed as Mo17, with the third Mo17 sample likely being a mislabeled RIL. Both validated IBM RILs and renamed RILs were retained for use in eQTL mapping and co-expression analysis. Only the validated RILs with original IBM names were used for leaf element correlation tests, which require sample identities to match across two

datasets. Only the two confirmed Mo17 samples were used in any analyses on the Mo17 parent genotype.

Bi-Parental Alignment Bias, Read Mapping, and Gene Quantification

Alignment bias is a complex problem for estimation of gene expression. The high level of genetic diversity in maize [35] is evident between the IBM population parents [27] and can introduce a bias in alignment depending on which parent's genome is used as a reference. If reads that harbor the reference allele preferentially align to the reference gene models, downstream expression analyses that seek to connect genetic background and expression may falsely associate the reference allele with an increase in gene expression. The B73 reference genome was the first maize reference genome released, and to date is the most widely used and highest quality reference available [26]. However, in order to reduce potential false positive associations, our alignment strategy must go beyond a standard alignment to B73 and incorporate the genetic variation between the two parents. Although the B73 genome was the first reference genome sequenced in maize and has been used as the sole reference in many studies, reference genomes of other maize genotypes have recently become available, including a reference for Mo17 [27]. Alignments of parent samples to both B73 and Mo17 references show a bias in read alignment towards the congenic reference. When using B73 as a reference, B73 samples aligned at rates of 83.9%, 82.4%, and 83.5%, while Mo17 samples aligned at rates of 73.5% and 72.4%. When using Mo17 as a reference, B73 samples aligned at rates of 68.5%, 67%, and 68.2%, while Mo17 samples aligned at rates of 73.5% and 72.3%. Alignment rates, as well as the bias, were greater using the B73 genome, likely due to differences in completeness of the two references.

In an attempt to address the alignment bias issue, we first considered aligning all samples to both the B73 reference transcriptome and Mo17 reference transcriptome and, within each

sample, choosing either the Mo17 reference or B73 reference transcript expression value for each gene. We assumed that if the exact same sample is aligned to both references, reads generated from a B73 region should align at a higher rate to the B73 reference than the Mo17 reference, with this difference being propagated to expression levels and the expression of a transcript from a B73 region having a higher value when aligned to B73 than Mo17. The opposite situation would be expected for Mo17 regions. Upon implementing this strategy, we discovered some unexpected challenges and found that our assumptions did not necessarily hold true. First, in order to choose between a B73 or Mo17-based gene expression value for each gene in each sample, we need to know which genes in B73 correspond to genes in the Mo17 annotation. Because there is no direct conversion system between the different nomenclatures and the coordinates of each genome are not on congruent scales, this requires comparative genomics querying to find allelic pairs between the two references. If a gene is present in only one of the parental genomes, or if it is not possible to determine an allelic pair, the strategy of aligning to both genomes does not apply. This issue limits the approach to only the genes that can be matched between the two genomes, and considering the diversity of the two parent genomes, could omit a significant amount of information from subsequent analyses.

Another issue with this approach also stems from the diversity of the two parent references. The assumption that reads from a B73 region of a RIL would align better to the B73 gene model becomes complicated when considering the entirety of each genome. The two references have varying copy numbers of genes, transposable elements, and other significant structural variation. These variations could bias alignment toward the reference with less copies of a particular gene. For example, if multiple copies of a gene are present in the B73 reference but only one copy exists in the Mo17 reference, reads that are actually from a B73 region of a

RIL may appear to align better to the Mo17 reference because that reference has less copies of the gene model as a search space. Similar, potentially even more complex, scenarios could arise from differences in transposable element insertion within genes. Discrepancies in quality and extent of gene annotation between the two references can also be problematic. If a gene pair has a more extensive annotation in one genome, for example more annotated transcripts in one genome or annotation of non-coding regions only in one genome, the differences make the B73 and Mo17 gene models non-equivalent search spaces for read alignment.

We also tested the idea of using SNPs to call breakpoints and generate custom references for each RIL through merging B73 and Mo17 sections. A major issue here is that the two genomes are not on equivalent scales, making it difficult to line up each genome at the same starting point, determine which specific blocks to pull out from each genome, and arrange the sections adjacently without including or excluding overlapping regions. The large gene order and chromosomal structural variation that occurs between B73 and Mo17 [27] only further confuses the matter. This strategy also still suffers from the issue of needing to find allelic pairs so a single gene can be quantified across all samples to be a trait for expression analysis.

The use of both references would be ideal if we could account for all of these potential complications stemming from the diversity of the two parent references and the inherent differences in quality and completeness of one reference versus another. Identifying all potential issues and developing a streamlined approach to utilizing both genomes is a largely uninvestigated area, particularly for species as diverse as maize and in bi-parental populations. Developing such an approach would be a project in its own right and is beyond the scope and goals of this investigation.

Read mapping strategy to minimize bias. Despite the inability to completely correct for bias, we still want to minimize mapping bias as much as possible in order to reduce false positive results in later analyses, all while maintaining a streamlined and executable alignment pipeline. Studies of allele specific expression have had to account for a similar reference bias problem and thus have developed some polymorphism-sensitive strategies to approach mapping [36–38]. The alignment issues in allele-specific experiments are very similar in nature to those seen here with a bi-parental population. WASP, a program developed for unbiased allele-specific read mapping, works to reduce allele-specific bias, has been previously tested to reduce false-positive eQTL more effectively than N-masked or personalized genome approaches, and easily integrates into our existing mapping pipeline [25]. With WASP, rather than working with two reference genomes, we can utilize the high quality and well-tested B73 reference along with an also high quality, dense set of Mo17 polymorphisms [28]. WASP integrates the Mo17 polymorphisms, 8.04 million SNP and insertion/deletion (indel) variants with 164 thousand coding sequence variants, by adding steps after the initial alignment to B73. WASP identifies reads that overlap SNPs and indels, discards reads that overlap indels and switches the allele(s) at reads overlapping SNPs to generate reads with all possible alternative allelic combinations, which are then remapped to the original reference. Reads overlapping SNPs that map to a different location of the genome when mapping any alternative allelic combinations are discarded.

To implement WASP, reads from 221 samples, with an average read count of 19.9 million reads per sample, were mapped to the B73 v4 transcriptome [26] with Tophat2, allowing for two mismatches, resulting in an average of 15.8 million mapped reads per sample. The average alignment rate among RILs for this first mapping was 78.5%. B73 parent samples mapped at rates of 83.9%, 82.4%, and 83.5%, while the Mo17 parent samples mapped at rates of

73.5% and 72.4%. Following the first mapping round, we input a set of mo17 polymorphisms to WASP and re-mapped SNP-overlapping reads using the same parameters. After remapping with WASP and discarding reads with mapping affected by allelic switches, an average of 14.6 million mapped reads per sample were retained. The average alignment rate among RILs was 72.5%. The B73 parent samples had mapping rates of 74.4%, 72.9%, and 74.0%, while the Mo17 samples had rate of 71.6% and 70.6%. Including WASP in the mapping pipeline reduced the largest mapping discrepancy among parent samples from 11.5% to 3.8% (Figure 2).

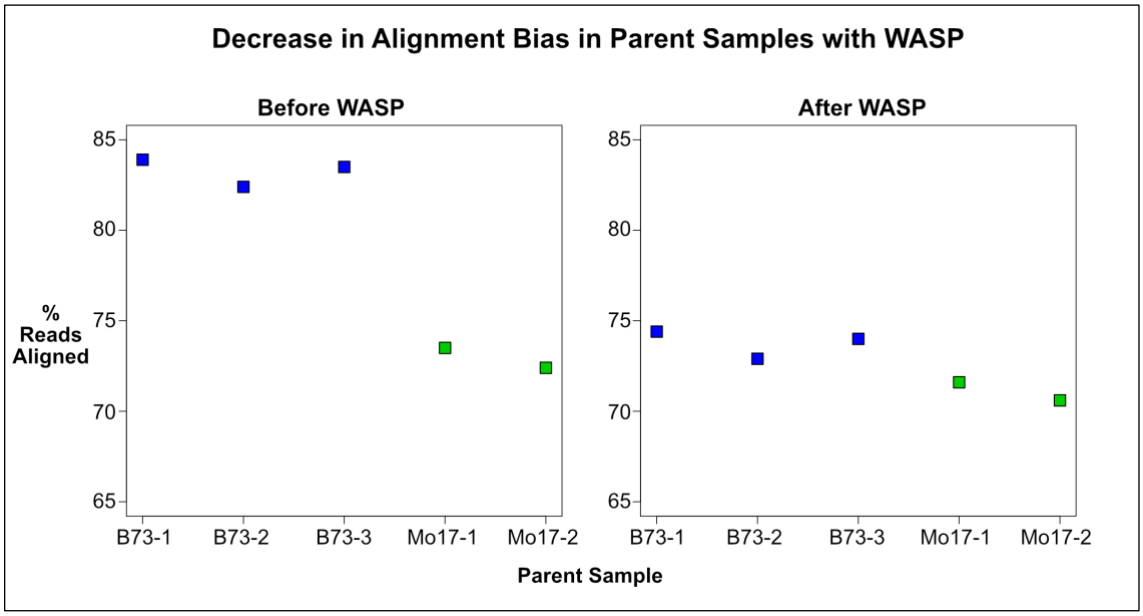


Fig 2. Parent Sample Alignment Before and After Bias Reduction. Percent reads aligned are shown for three B73 parent samples and two Mo17 parent samples with and without inclusion of SNP-based bias correction. The left panel shows alignment rates from alignment to the B73 v4 reference. The right panel shows alignment rates from alignment to B73 v4 plus WASP.

While addition of WASP is not a perfect solution as it does not consider large structural variation and it reduces overall alignment by discarding some reads that overlap polymorphisms, it does reduce bias as is shown in the reduction of bias in parent alignment rates. The alignment rate decrease of 6% among RILs with the application of WASP is a trade-off in order to avoid

including a large proportion of the reads that would map with allelic bias and could introduce false positives in expression analyses.

Gene quantification. Gene expression was quantified in 215 RILs using Cufflinks2 [39] and the alignments from the Tophat2 plus WASP pipeline. The B73 v4 transcriptome was used to guide gene quantification and limit quantification only to annotated genes. Gene expression was estimated and scaled using RPKM (read counts per kilobase of exon per million mapped reads). 38,629 genes were quantified in at least one accession. Filtering out genes expressed in less than 80% lines and genes with an expression mean of less than 0.5 RPKM resulted in a set of 26,440 expressed genes. Of these filtered expressed genes, 93.6% had mean expression across all lines greater than 1 RPKM, 54.2% had mean expression over 10 RPKM, 15.9% had a mean over 50 RPKM, and 0.07% had mean over 100 RPKM. 17,303 genes were expressed across all samples. The number of genes expressed per sample using a cutoff of 1 RPKM ranged from 20,355 to 24,375, with a mean of 23,488 genes expressed per sample, while the number of genes expressed per sample above 10 RPKM ranged from 11,347 to 14,535, with a mean of 13,778 genes expressed per sample (Fig S1).

Genetic Control of Gene Expression in the Maize Root

Global eQTL mapping. Expression QTL mapping on 26,440 genes measured in 215 RILs was performed using the program Matrix eQTL [29]. The linear model setting was used in matrix eQTL, with p-value thresholds set at 2e-10 and 1e-10 for cis and trans eQTL, respectively, and cis-eQTL set as associations across a distance of less than 1Mb, resulting in an initial result of 25,629 cis-eQTL and 114,116 trans-eQTL that passed p-value and FDR cutoffs. Because of the extent of linkage disequilibrium in this population, the 139,745 initially reported eQTL actually comprise far fewer eQTL because linked SNPs will each return an association

with a gene. To assess results, we only want to consider windows of linked SNPs that reflect the level of resolution achievable with this genetic map, which is not high enough to get to the level of individual SNPs.

To collapse linked SNP windows, hierarchical clustering was performed on each gene with an eQTL (eGene). All significant SNPs (eSNPs) were considered for each eGene, with eSNPs that clustered together based on a correlation cutoff of 0.75 being merged into eSNP windows. With this removal of linked SNP redundancy, we retained a final set of 19,320 eQTL consisting of eSNP windows associated with eGenes (Figure 3A). This final set includes 7,160 unique eGenes. 15,407 eQTL have eSNP and eGene on the same chromosome, while 3,913 have associations between different chromosomes. Of the eQTL with eSNP and eGene on the same chromosome, 6,049 eQTL had a distance of 1 Mb between eSNP and eGene, 10,922 had a distance of 5 Mb, and 12,497 had a distance of 10 Mb. 2,210 eQTL had eSNP and eGene on the same chromosome but separated by greater than 15 Mb. Using an arbitrary definition of cis-eQTL being those where the eSNP and eGene are on the same chromosome and within 10 Mb of each other, 12,497 cis-eQTL were mapped, containing of 5,889 unique eGenes. Of these cis-eQTL, 7,197 showed an increase in expression with the B73 allele, and 5,300 showed an effect in the Mo17 direction. Of the cis-eQTL that occurred across windows of 1Mb or smaller, 3,426 had an effect in the B73 direction and 2,623 had an effect in the Mo17 direction. Using a definition of trans-eQTL as eQTL where eSNP and eGene are either on different chromosomes or over 15 Mb apart on the same chromosome, we detected 6,128 trans-eQTL, containing 2,416 unique genes. 2,118 trans-eQTL showed increased expression with the B73 allele, while 4,005 showed increased expression with the Mo17 allele.

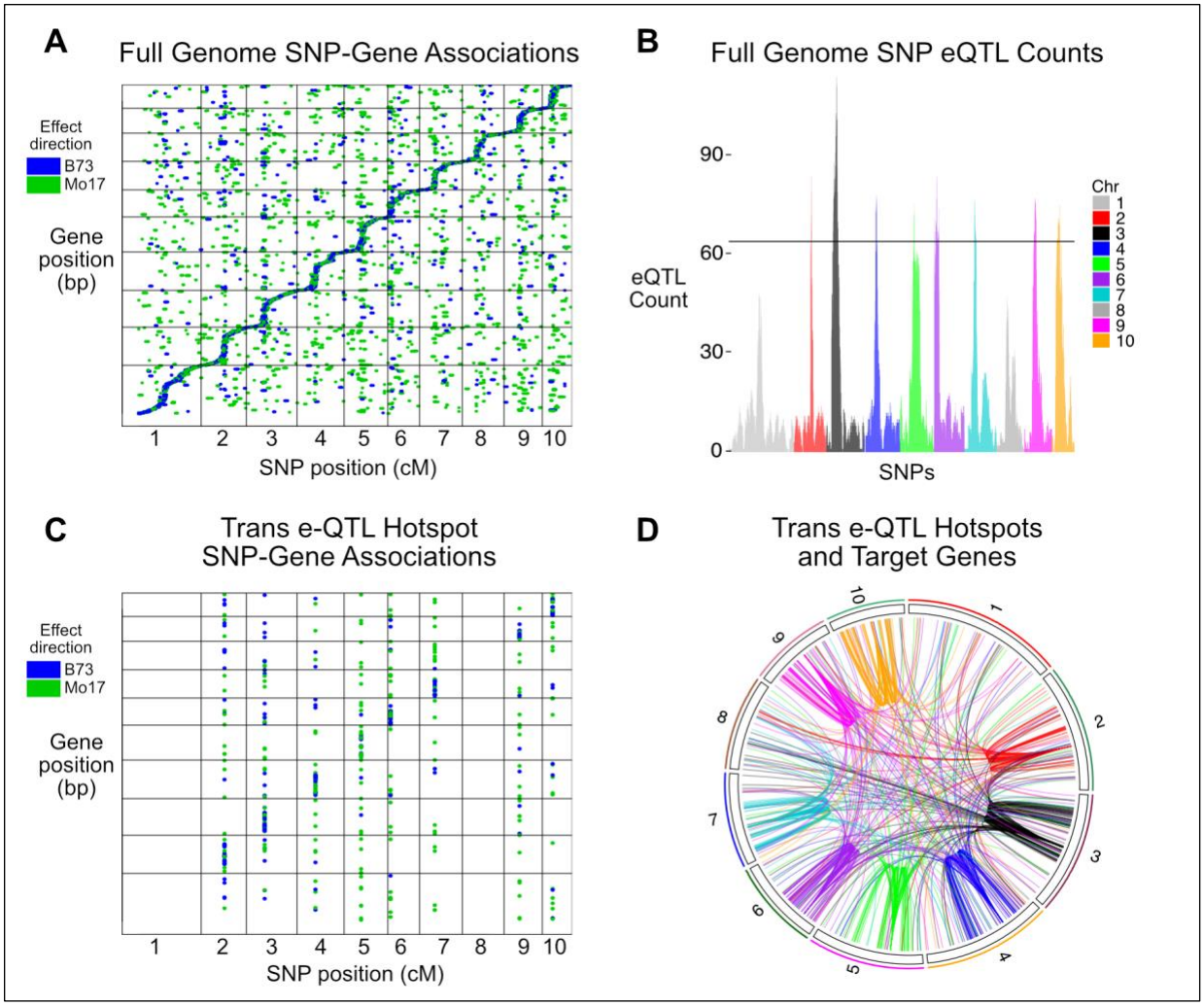


Fig 3. Expression QTL and Trans-eQTL Hotspots. (A) Points represent most significant eSNP of eSNP window and target eGene. SNP position is represented on the x-axis in cM and gene positions are represented on the y-axis in base pairs. Points are colored by effect direction, with green representing increased expression with the Mo17 allele and blue showing increased expression with the B73 allele. (B) Barplot of number of eQTL per SNP. Horizontal line indicates 95th percentile of eQTL counts. (C) Graph is the same layout as (A) but shows only eSNPs and eGenes of trans-eQTL hotspots. (D) Circos plot showing eQTL of trans-eQTL hotspots. Links connect top representative SNP from each hotspot to the hotspot gene targets. Links are colored by hotspot.

The greater occurrence of cis-eQTL where the B73 allele increases expression may reflect remaining alignment bias that was not removed with our approach. Trans-eQTL could also be impacted by alignment bias due to larger structural variations or paralogs (Fig 4), which would not have been accounted for in our SNP-sensitive pipeline. In a case where both parents

share a gene but only one parent has a distally located paralog of that gene, it could appear as though the expression of the gene is associated with the allele at the paralog location, even if the expression of the gene is consistent across RILs. A check for false positives could be conducted by looking at parent sample alignments to both B73 and Mo17 genomes for this specific set of genes, although the implementation, as discussed above, has its own challenges. At this time, a list of syntelogs or allelic pairs between the B73 v4 genome and the Mo17 genome is not available. However, we do have such a list linking the B73 v3 reference and an older version of the Mo17 reference. Of the 2,416 genes with trans-eQTL, only 998 genes have a paired Mo17 gene and could be looked into further using parent sample alignments. When parent samples were aligned to both the B73 and Mo17 versions of these 998 genes, 822 genes have same effect direction regardless of which genome is used for alignment, while 176 have a different effect direction depending on reference. This provides a very limited glimpse into the potential level of false positives caused by mapping bias. The nature and magnitude of effect differences could possibly be further dissected, however, given the vast number of scenarios that can lead to alignment differences depending on reference and the ability to evaluate only a fraction of these genes, such an exercise is unlikely to produce definitive global estimates of alignment-induced false positives. Evaluating eQTL that are of interest for further investigation on a case-by-case basis may be the most effective approach at this juncture until a new comparison can be made between the B73 v4 and Mo17 references.

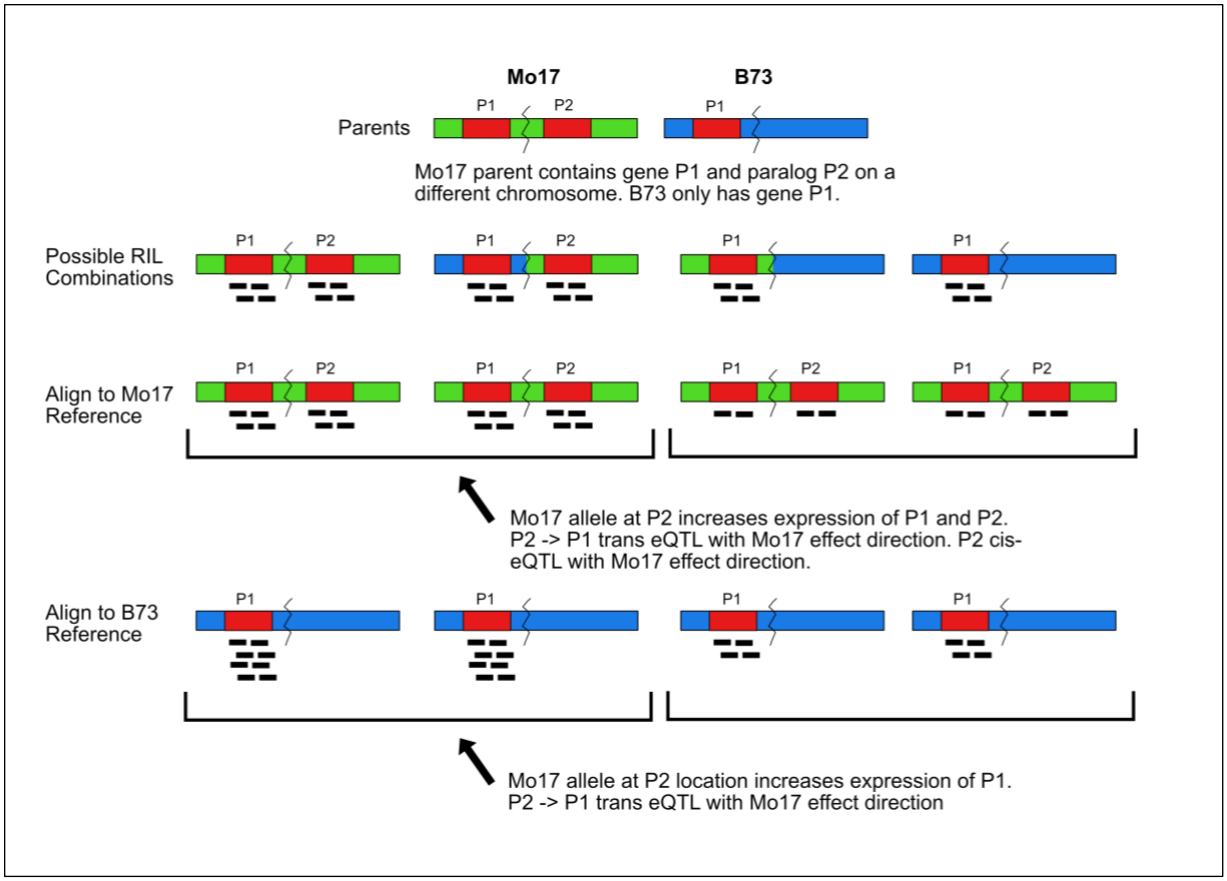


Fig 4. Differential Paralogs and False-Positive Trans-eQTL. Schematic of a scenario in which only one parent having an additional gene copy causes a false positive trans-eQTL. Gene P1 has a paralog, P2, located on a different chromosome, in Mo17. Gene P1 does not have a paralog in B73. All possible RIL combinations are shown, with P1 and P2 each producing four reads for this visual demonstration. When RILs are aligned to either reference genome, having the Mo17 allele at the P2 location appears to increase expression of P1, although the expression of P1 is actually consistent across RILs.

Trans-eQTL Hotspots. Trans-eQTL were evaluated to detect any trans-eSNPs with exceptionally high numbers of target genes that could be considered trans-eQTL hotspots. Trans-eQTL hotspots were identified by counting the number of trans-eQTL (target gene on a different chromosome or over 15Mb away on the same chromosome) per SNP (Fig 3B). All SNPs were considered and given a trans-eQTL count when present in a trans-eQTL window. SNPs with counts exceeding the 95th percentile of all SNP counts were considered hotspots and SNPs correlated with a Pearson's correlation coefficient above 0.9 were merged into a single hotspot.

Eight trans-eQTL hotspots were identified, with one on each chromosome other than chromosomes 1 and 8 (Fig 3D & E). The number of genes targeted by hotspots ranged from 78 to 151. These gene targets include genes more local to the hotspot, near the 15 Mb cutoff, farther ranging genes on the same chromosome, and genes on different chromosomes (Table 1).

Table 1. Trans-eQTL Hotspot SNP and Gene Counts.

Hotspot	Num SNPs	Num Genes	Genes on Same Chr (> 15 Mb)	Genes on Diff Chr
Hotspot Chr2	2	89	50	39
Hotspot Chr3	7	151	99	52
Hotspot Chr4	8	91	64	27
Hotspot Chr5	5	80	46	34
Hotspot Chr6	11	98	60	38
Hotspot Chr7	10	78	48	30
Hotspot Chr9	12	82	46	36
Hotspot Chr10	8	95	67	28

The genes associated with trans-eQTL hotspots were tested for GO term enrichment using a hypergeometric calculation and a multiple testing correction for multiple comparisons. Genes targeted by 7 of the 8 hotspots were enriched for one or more GO terms (Table S5). The most significant GO term enrichment for each cluster is listed in Table 2. The values shown in Table 2 reflect the hypergeometric calculation as follows: given the number of genes in a hotspot ("target term size") out of all the genes in a genome, the p-value is the probability of finding "num com" number of those genes when you sample the number of genes with that GO term ("source term size") out of out of all the genes in the genome ("num univ"). Multiple correction is applied to account for "terms tested" which reflects the number of GO terms tested for a particular hotspot. These enrichments are by no means exhaustive owing to the limited functional annotation of the maize genome, but can offer a high-level view of coordinated functions of genes targeted by trans-eQTL hotspots.

Table 2. Trans-eQTL Hotspot Top GO Term Enrichments.

Hotspot	GO Term	ID	p-val	terms tested	num com	num univ	source term size	target term size	num terms
Hotspot Chr2	cohesin complex	GO:0008278	2.83E-06	367	4	39,324	43	89	11,909
Hotspot Chr3	Lys63-specific deubiquitinase activity	GO:0061578	1.04E-09	789	4	39,324	5	151	11,909
Hotspot Chr4	phenylalanine ammonia-lyase activity	GO:0045548	1.01E-07	581	4	39,324	19	91	11,909
Hotspot Chr5	histone acetyl-transferase binding	GO:0035035	2.37E-10	406	4	39,324	6	80	11,909
Hotspot Chr6	DNA packaging	GO:0006323	3.55E-08	498	5	39,324	37	98	11,909
Hotspot Chr9	acireductone dioxygenase [iron(II)-requiring] activity	GO:0010309	7.27E-07	555	3	39,324	9	82	11,909
Hotspot Chr10	shikimate 3-dehydrogenase (NADP+) activity	GO:0004764	4.75E-07	571	3	39,324	7	95	11,909

Co-Expression of Genes in the Maize Root

Co-Expression Analysis. Co-expression analysis was performed using the Camoco framework [7], which computes Pearson correlation coefficients between expressed genes and implements a MCL (Markov Cluster) algorithm to find co-expressed gene clusters. A raw expression file of 38,639 genes that had any level of expression across any of 221 samples (RILs and parent samples) was filtered to remove genes missing more than 20% of data and require a minimum expression level above 0.01 RPKM before co-expression calculations. The co-expression network derived with Camoco contained 24,354 genes (63% of total) and 250 MCL

clusters with a size greater than 10 (Fig 5). The level of gene ontology (GO) term co-expression was 15.5-fold higher than expected by chance (Fig S2).

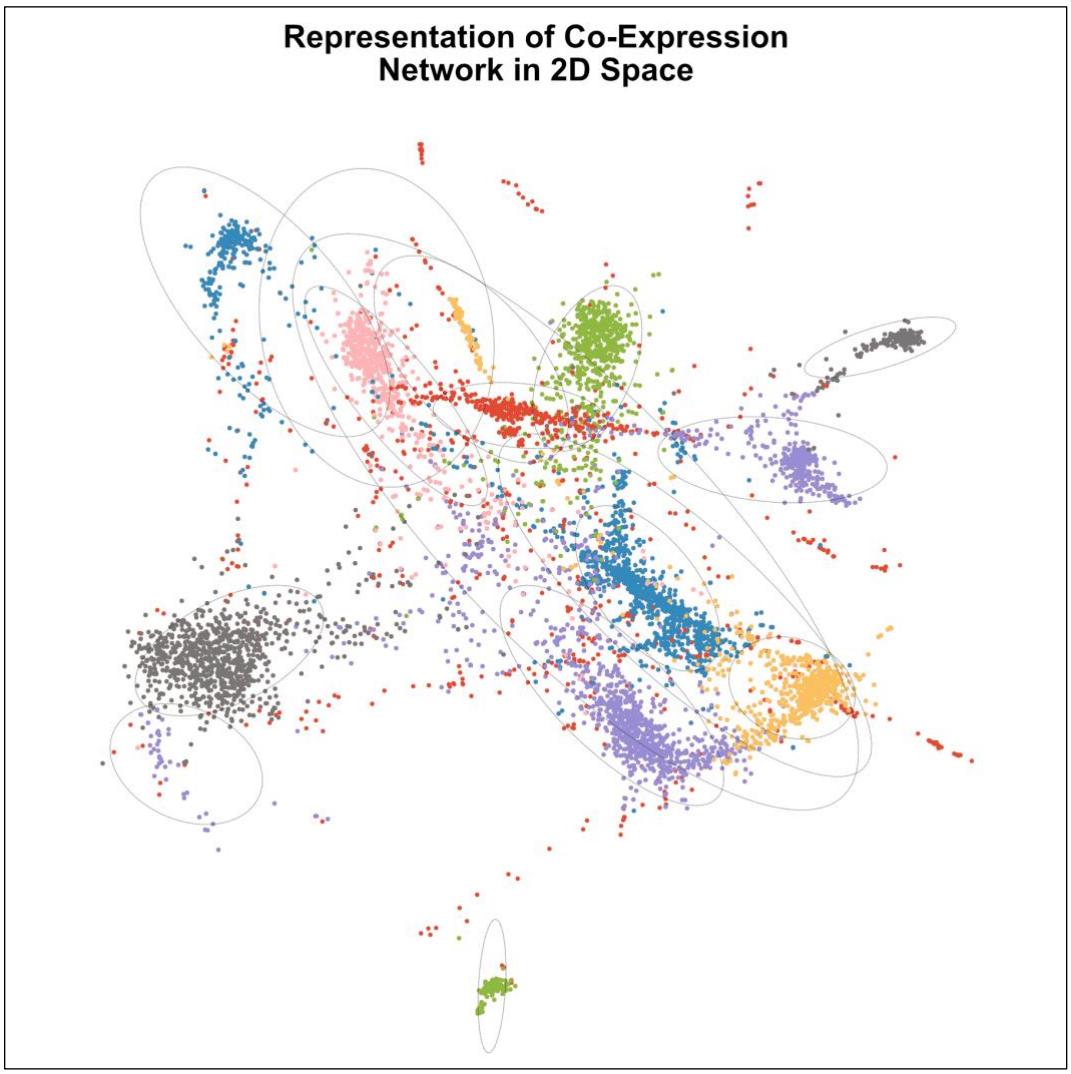


Fig 5. Co-Expression Network. Representation of clusters generated using a fast force directed layout algorithm. Sets of nodes with higher co-expression are closer in 2D space in the plot, but placement is not quantitative. Colors are used to differentiate between MCL clusters. Circles are shown around MCL clusters with greater than 100 genes. The circles are computed by calculating the first two principle components using the gene coordinates for each MCL cluster to produce ellipse parameters.

Trans-eQTL MCL Cluster Enrichments. Target gene lists of the 8 trans-eQTL hotspots were tested for over-representation within gene lists for co-expressed MCL clusters. The hypergeometric calculation with multiple testing correction used for GO term enrichment was also used for MCL enrichment tests. As expected for co-regulated sets of genes, all sets of trans-eQTL gene targets were significantly enriched for co-expression, with the number of MCL cluster enrichments per hotspot ranging from four to seven (Table S6). We further explored the four MCL cluster enrichments for genes associated with the trans-eQTL hotspot on chromosome 2 (Table 3). 70 of the 89 chromosome 2 hotspot gene targets belong to one of four different clusters, with genes in the same cluster showing similar expression patterns and the same effect direction (Fig 6).

Table 3. Chromosome 2 Trans-eQTL Hotspot MCL Enrichments.

Hotspot	**MCL Cluster**	**p-val**	**terms tested**	**num com**	**num univ**	**source term size**	**target term size**	**num terms**
HotspotChr2	MCL28	1.34E-78	9	36	24,430	53	89	4,042
HotspotChr2	MCL70	4.59E-34	9	16	24,430	25	89	4,042
HotspotChr2	MCL85	5.39E-22	9	11	24,430	22	89	4,042
HotspotChr2	MCL76	2.20E-12	9	7	24,430	24	89	4,042

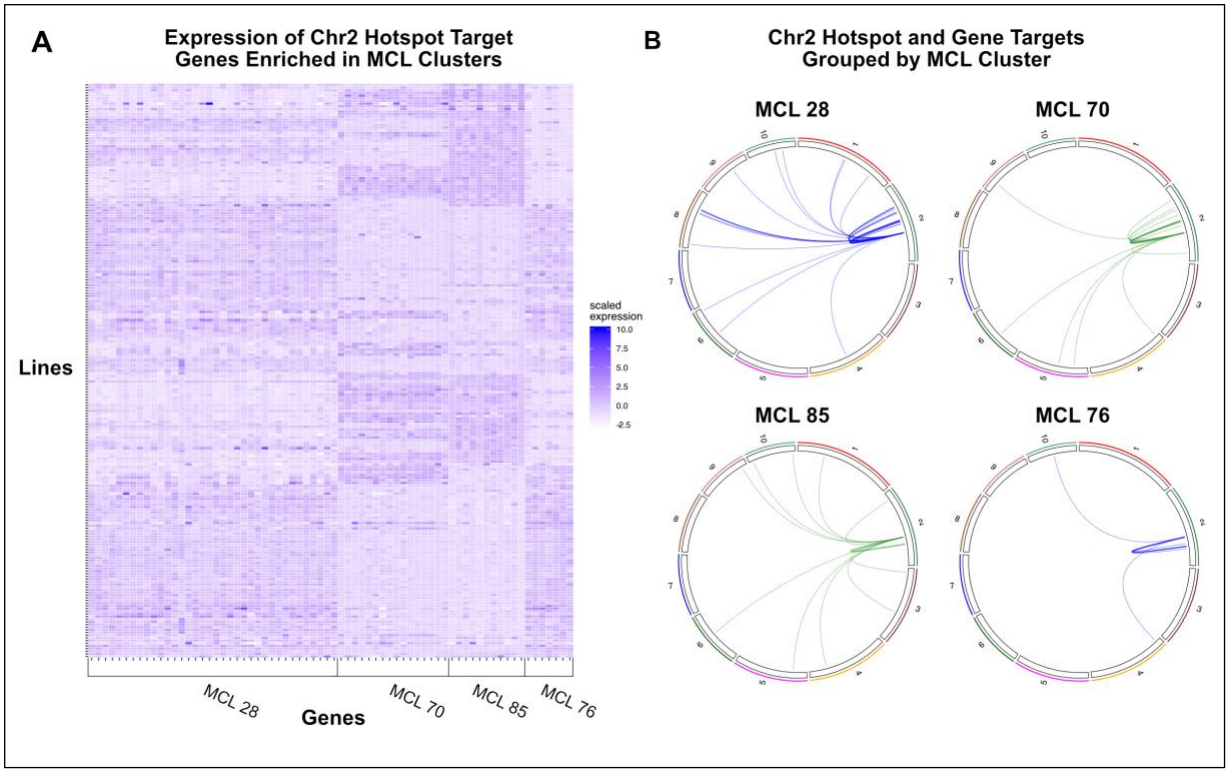


Fig 6. Chromosome 2 Hotspot MCL Enrichment. (A) Heatmap of expression of chr2 hotspot target genes within MCL clusters, shown across all RILs. Genes are grouped by their MCL cluster. (B) Circos plots of chr2 hotspot eQTL. Gene targets in different MCL clusters are shown in different plots, with links colored by effect direction. Blue indicates B73 direction and green indicates Mo17 direction.

Linking Root Gene Expression and Leaf Ionome

Correlations were tested for between gene expression in the roots and element profiles in the leaves of the IBM RILs. Testing for these correlations is a way to survey genes expressed in the root for a potential connection with leaf element accumulation. Only the 146 RILs that had validated IBM sample identity from SNP comparison were used for the correlation tests. The Pearson's correlation coefficient was calculated between each of the 20 elements measured in the leaf and the root-based expression of 26,386 genes with expression in at least 20% of lines and with a mean expression value of at least 0.5 RPKM. The top ten correlated genes were recorded for each element (Table S7).

Genes that were within the top ten correlated for an element were evaluated with respect to QTL detected for that element in the leaf. 10 of the 13 leaf QTL have at least one top ten correlated gene within 10 Mb of the QTL location (Table 4). Many of these genes have cis-eQTL, as might be expected to occur transitively given that loci in the gene region are associated with a particular element and the gene is also correlated with element concentration. In this sense, these cis-eQTL can be thought of as markers for the element traits. These correlated genes near leaf QTL provide potential candidate genes for element QTL, and some may be worthwhile of further investigation through the use of other data types or additional experiments. However, these lists should be considered in the context of the of genetic resolution in the IBM population and the increased likelihood of co-expression among neighboring genes. Genes that are not the causal genes for a QTL but are within a region exhibiting low recombination with the causal gene are likely to appear in such lists solely due to their genetic location rather than because of a functional connection.

Table 4. Leaf Element QTL and Overlapping Element-Correlated Genes

Leaf El QTL	Gene	El Cor	Chr	Pos (bp)	Gene Function	Cis eQTL?	eQTL Effect
Al 5@150.9 60,565,262	Zm00001d014795	0.37	5	63,482,860	Unknown	Yes	B73
	Zm00001d014726	0.37	5	60,843,523	Mitogen-activated protein kinase 17	Yes	B73
As 3@346.4 215,434,459	Zm00001d044146	-0.60	3	220,480,167	cytochrome P450 family 72 subfamily A polypeptide 8	No	NA
Ca 10@245.5 148,723,309	Zm00001d026628	-0.43	10	149,177,141	Unknown	No	NA
Cd111 2@214.6 168,260,178	Zm00001d005195	-0.65	2	163,095,556	RING/U-box superfamily protein	Yes	B73
	Zm00001d005429	-0.62	2	173,709,958	COP1-interacting protein-related	Yes	B73
	Zm00001d005231	-0.59	2	164,761,161	ADP-ribosylation factor A1F	Yes	B73
	Zm00001d005489	0.56	2	175,795,127	D-isomer specific 2-hydroxyacid dehydrogenase family protein	Yes	Mo17

	Zm00001d005295	-0.55	2	168,137,133	DNA-directed RNA polymerases II IV and V subunit 3	Yes	B73
K39 3@352.6 217,198,934	Zm00001d044159	0.59	3	220,748,190	Unknown	No	NA
Mo98 1@378.0 248,800,963	Zm00001d033080	0.66	1	249,851,631	QWRF motif-containing protein 3	Yes	Mo17
	Zm00001d033111	-0.62	1	250,442,305	Putative lysine decarboxylase family protein	Yes	B73
	Zm00001d032968	-0.59	1	244,998,348	Tetrapyrrole (Corrin/Porphyrin) Methylases	Yes	B73
	Zm00001d033226	-0.55	1	255,204,711	Unknown	Yes	B73
Na23 4@196.9 162,739,209	Zm00001d051525	0.56	4	161,345,010	Oligopeptide transporter 4	No	NA
Ni60 9@7.7 1,840,217	Zm00001d044768	0.44	9	1,932,258	Protein NRT1/ PTR FAMILY 5.8	Yes	B73
S34 1@416.2 274,044,141	Zm00001d033818	0.35	1	274,726,910	Transmembrane and coiled-coil domains protein 1	Yes	B73
	Zm00001d033575	-0.33	1	267,540,332	DUF1336 domain containing protein expressed	Yes	Mo17
	Zm00001d033750	0.32	1	272,770,260	Threonine dehydratase biosynthetic chloroplastic	Yes	B73
Zn66 1@401.0 262,566,563	Zm00001d033584	0.34	1	267,873,372	Unknown	Yes	B73
	Zm00001d033590	-0.32	1	267,998,422	Ribosomal L18p/L5e family protein	Yes	Mo17
	Zm00001d033575	-0.31	1	267,540,332	DUF1336 domain containing protein expressed	Yes	Mo17
	Zm00001d033446	0.30	1	262,975,467	Zinc transporter 7	Yes	B73
	Zm00001d033307	0.29	1	258,331,851	Outer arm dynein light chain 1 protein	Yes	B73
	Zm00001d033469	0.29	1	263,524,861	Ferredoxin%253B Putative ferredoxin	Yes	B73
	Zm00001d033189	0.29	1	253,284,685	Unknown	No	NA

Genes correlated with leaf element concentration were also examined for trans-associations with leaf element QTL. These associations were less common, only present for two of the 13 leaf QTL (Table 5), yet represent possibly more interesting functional connections than the cis-associations. Rather than connections between a locus already associated with element accumulation and genes near that locus, these trans-associations link the locus associated with element accumulation to genes not near the known locus, bringing a new genetic region into the

picture.

A large effect leaf Cd QTL (which was also previously detected in the seed) is on chromosome 2 around 214.6 cM (equivalent to 168 Mbp) and SNPs in that region are trans-eQTL for all five genes in the top ten list of correlated expression with leaf Cd content that are not on chromosome 2 (Fig 7). These include genes on chromosomes 4, 5, 6, and 10. The SNP ranges on chromosome 2, as shown in Table 5, reflect the genetic resolution of the QTL/eQTL region, which is relatively broad. However, the eGene targets offer gene-level resolution of other regions in the genome that interact with the Cd QTL and provide interesting links for further investigation of Cd accumulation. The Cd QTL region on chromosome 2 collocates with the chromosome 2 trans-eQTL hotspot, with links to these top Cd correlated genes as well as to other regions of the genome. In addition to the Cd-correlated genes on chromosomes 4, 5 and 6, the hotspot targets four other genes on chromosome 4, two other genes on chromosome 5, and four other genes on chromosome 6. On chromosome 10, the hotspot targets three other genes aside from the two Cd-correlated genes. The five genes correlated with leaf Cd are mainly targeted by the chromosome 2 leaf Cd QTL/trans-eQTL hotspot region, with only one additional eQTL on chromosome 7 associated with one of the chromosome 10 genes. Of the five other genes correlated with leaf Cd that are in the chromosome 2 Cd QTL region, only one is associated with SNPs outside of the region through a trans-eQTL with the eSNP located closer to the beginning of chromosome 2, over 80 Mb away from the Cd QTL region (Fig 7).

The leaf Zn QTL on chromosome 1 is also connected with a gene elsewhere in the genome through a trans-eQTL located within the Zn QTL region that targets a gene on chromosome 5. The chromosome 5 gene was found to have the highest root expression correlation with leaf Zn of all genes outside of the Zn QTL region.

Table 5. Leaf Element QTL and eQTL with Element-Correlated Gene Targets

Leaf QTL	eQTL Chr	eQTL Pos	eGene	Gene Chr	Gene Pos	Gene - El Cor	Function	eQTL Effect Direction
Cadmium 2@214.6 168,260,178	2	153,447,098 - 172,156,107	Zm00001d014345	5	42,127,512	0.68	C2H2-like zinc finger protein	Mo17
	2	153,447,098 - 191,280,878	Zm00001d023657	10	13,783,362	0.6	Unknown	Mo17
	2	153,447,098 - 172,156,107	Zm00001d036628	6	95,592,329	0.58	Single-stranded nucleic acid binding R3H protein	Mo17
	2	153,447,098 - 172,156,107	Zm00001d052443	4	190,127,612	0.56	Unknown	Mo17
	2	153,447,098 - 172,156,107	Zm00001d024560	10	78,246,604	-0.55	RING/U-box superfamily protein	B73
Zinc 1@401.0 262,566,563	1	264,213,949 - 277,355,310	Zm00001d017634	5	202,684,054	-0.32	DUF936 family protein	Mo17

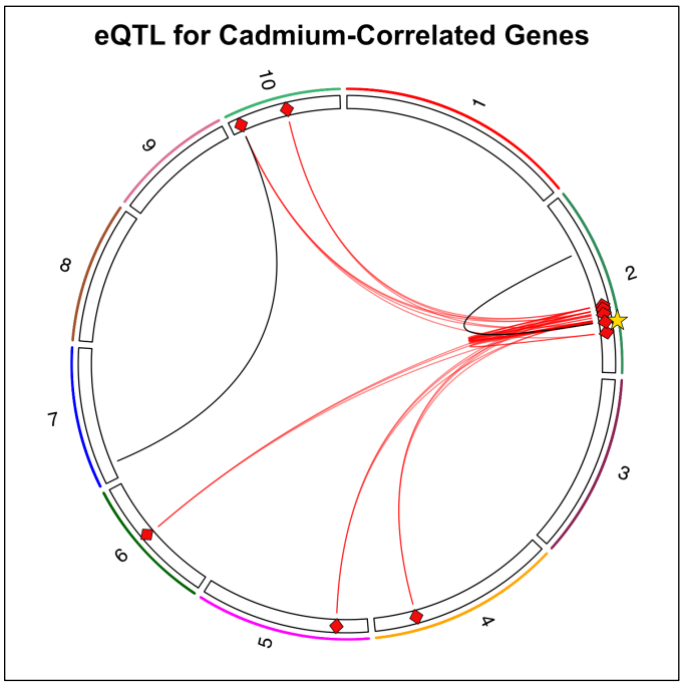


Fig 7. Trans-eQTL and Cadmium-Correlated Genes. All significant eQTL across the genome for the top 10 genes correlated with Cd (represented with red diamonds) are shown, with red lines indicating eSNPs in the Chr2 hotspot region (represented with a star) and black lines indicating eSNPs elsewhere in the genome.

DISCUSSION

In this study we associated genetic markers with leaf element phenotypes through QTL mapping, linked genetic markers to root gene expression through eQTL mapping, and related root gene expression with the leaf ionome through correlation analysis. By dissecting these relationships, we were able to develop candidate gene lists within leaf and seed ionome QTL regions. We also gained insight into genetic regulatory networks that involve previously mapped loci, identifying trans-eQTL that overlap with ionome QTL and connect to expression of genes elsewhere in the genome.

We found several cases where the genes most correlated with element concentration collocate with element QTL. While these genes certainly require further testing to confirm and describe associations with element regulation, they provide starting candidate gene lists of a reasonable size compared to the number of genes typically in a QTL confidence interval. For example, a zinc QTL on chromosome 1 detected in both the seed and leaf overlaps the genes with the top expression correlations with zinc, one of which is annotated as Zinc transporter 7. In this case, we have a locus detected in multiple tissues, across multiple environments, with a promising gene candidate found by analyzing gene expression in an additional tissue. These lines of evidence warrant additional exploration of this gene, which has an Arabidopsis ortholog belonging to a list of known ionome genes, genes that have been associated with particular ionomic phenotypes. Zinc transporter 7 is a member of a larger family of zinc transporters, ZIP proteins. Studies in Arabidopsis, rice, and maize have linked ZIP proteins to not only Zn uptake but also Fe transport and storage, metal homeostasis, and salinity and drought tolerance [40–43].

Root gene expression data additionally provided insight into a cadmium QTL on chromosome 2, also detected in the leaf and seed in multiple environments. The Cd QTL region

coincides with a trans-eQTL hotspot regulating a large set of genes that can be broken down into co-expressed modules. Notably, the most cadmium-correlated genes of all genes expressed in the root either collocate with the Cd QTL or are trans-eQTL gene targets of eQTL in the Cd QTL region. These results not only provide information that may narrow down the Cd QTL interval, but also suggest other genes that could be involved in a gene regulatory network with some impact on Cd accumulation. This work is an additional step forward in finding genes that control the complex process of heavy metal uptake and storage. Further investigation of these regulatory networks in the root could advance the effort of developing crop variants that can be grown in areas with high heavy metal concentrations without storing toxic levels of heavy metals in food source tissues.

The ionome shows a strong interaction with environment, and thus environmental effects may prevent us from characterizing previously mapped QTL from field environments with new data from a greenhouse environment. However, the consistent use of the IBM population across studies has allowed us to use greenhouse-generated data to further understand certain QTL with large effects across different environments. Genes at Zn and Cd QTL, as well as genes that have trans-eQTL associations overlapping the Cd locus, are promising candidates for future work on element regulation. While our previous QTL mapping studies on the ionome detected many loci, the number of genes within a QTL interval is generally exceedingly large and cannot provide evidence to justify further use of resources to study particular genes within the interval. Using root-based RNA-sequencing, we measured expression at the gene level in a tissue known to be highly determinant of the whole-plant ionome, adding a unique and pertinent layer of support for characterizing QTL.

REFERENCES

1. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, Danku J et al. Variation in Molybdenum Content Across Broadly Distributed Populations of *Arabidopsis thaliana* Is Controlled by a Mitochondrial Molybdenum Transporter (*MOT1*). PLoS Genet. 2008;4: e1000004.
2. Baxter I. Ionomics: The functional genomics of elements. Briefings in functional genomics. 2010;9: 149-156.
3. Borghi M, Rus A, Salt DE. Loss-of-function of Constitutive Expresser of Pathogenesis Related Genes5 affects potassium homeostasis in Arabidopsis thaliana. PLoS One. 2011;6: e26360.
4. Holloway B, Luck S, Beatty M, Rafalski J-A, Li B. Genome-wide expression quantitative trait loci (eQTL) analysis in maize. BMC Genomics. 2011;12: 1-14.
5. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Phil Trans R Soc B. 2013;368: 20120362.
6. Schaefer RJ, Michno J-M, Myers CL. Unraveling gene function in agricultural species using gene co-expression networks. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms. 2016
7. Schaefer R, Michno J-M, Jeffers J, Hoekenga OA, Dilkes BP, Baxter IR et al. Integrating co-expression networks with GWAS to prioritize causal genes in maize. The Plant Cell. 2018tpc.00299.2018.
8. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013;45: 43-50.
9. Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z et al. RNA sequencing reveals the complex regulatory network in the maize kernel. Nature communications. 2013;4:
10. Veley KM, Berry JC, Fentress SJ, Schachtman DP, Baxter I, Bart R. High-throughput profiling and analysis of plant responses over time to abiotic stress. Plant Direct. 2017;1: e00023.
11. Ziegler G, Terauchi A, Becker A, Armstrong P, Hudson K, Baxter I. Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. The Plant Genome. 2013;6:
12. Asaro A, Ziegler G, Ziyomo C, Hoekenga OA, Dilkes BP, Baxter I. The Interaction of Genotype and Environment Determines Variation in the Maize Kernel Ionome. G3: Genes Genomes Genetics. 2016;6: 4175-4183.
13. Davies L, Gather U. The Identification of Multiple Outliers. Journal of the American Statistical Association. 1993;88: 782-792.
14. Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64: 641-656.
15. Andrews S. FastQC: A quality control tool for high throughput sequence data. Reference Source. 2010
16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014btu170.
17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids

research. 2013;41: D590-D596.
18. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9: 357-359.
19. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science. 2009;326: 1112.
20. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14: R36.
21. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 2009;25: 2283-2285.
22. Ganal MW, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A et al. A Large Maize (Zea mays L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. PLOS ONE. 2011;6: e28334.
23. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM et al. dbSNP: the NCBI database of genetic variation. Nucleic acids research. 2001;29: 308-311.
24. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156-2158.
25. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nature methods. 2015;12: 1061-1063.
26. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546: 524-524.
27. Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nature Genetics. 2018;50: 1289-1295.
28. Zhou P. Maize Mo17 SNPs. Retrieved from the Data Repository for the University of Minnesota. https://doi.org/10.13020/D6T38X:
29. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28: 1353-1358.
30. Byng MC, Whittaker JC, Cuthbert AP, Mathew CG, Lewis CM. SNP Subset Selection for Genetic Association Studies. Annals of Human Genetics. 2003;67: 543-556.
31. Wimalanathan K, Friedberg I, Andorf CM, Lawrence-Dill CJ. Maize GO Annotation: Methods, Evaluation, and Review (maize-GAMER). Plant Direct. 2018;2: e00052.
32. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. PLOS ONE. 2014;9: e98679.
33. Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D et al. Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant molecular biology. 2002;48: 453-461.
34. Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics. 1994;138: 963-971.
35. Buckler ES, Gaut BS, McMullen MD. Molecular and functional diversity of maize. Current opinion in plant biology. 2006;9: 172-176.
36. Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y et al. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data.

Bioinformatics. 2009;25: 3207-3212.
37. Stevenson KR, Coolon JD, Wittkopp PJ. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. BMC Genomics. 2013;14: 536.
38. Ahn J, Xiao X. RASER: reads aligner for SNPs and editing sites of RNA. Bioinformatics. 2015;31: 3906-3913.
39. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012;7: 562-578.
40. Oyiga BC, Sharma RC, Baum M, Ogbonnaya FC, Leon J, Ballvora A. Allelic variations and differential expressions detected at quantitative trait loci for salt stress tolerance in wheat. Plant, cell & environment. 2018;41: 919-935.
41. Li S, Zhou X, Zhao Y, Li H, Liu Y, Zhu L et al. Constitutive expression of the ZmZIP7 in Arabidopsis alters metal homeostasis and increases Fe and Zn content. Plant Physiology and Biochemistry. 2016;106: 1-10.
42. Milner MJ, Seamon J, Craft E, Kochian LV. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. Journal of Experimental Botany. 2013;64: 369-381.
43. Mondal T, Ahmad Ganie S, Rana M, Sharma T. Genome-wide Analysis of Zinc Transporter Genes of Maize (Zea mays). 2014

SUPPORTING INFORMATION

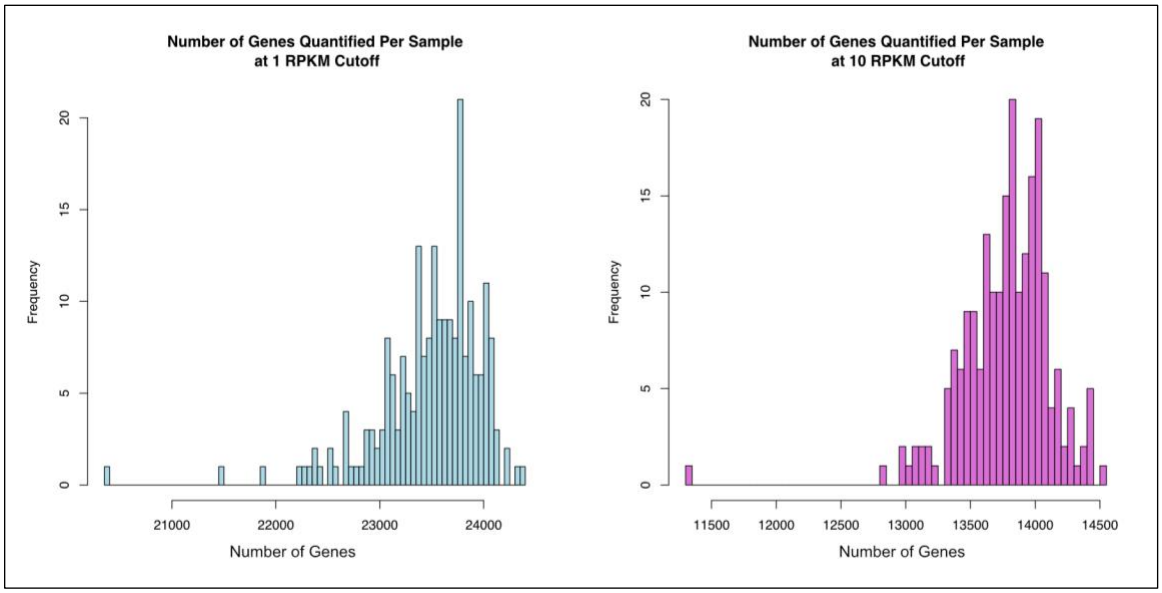


Fig S1. Genes Expressed per Sample. Distributions of number of genes quantified across samples are shown for 1 RPKM and 10 RPKM cutoffs.

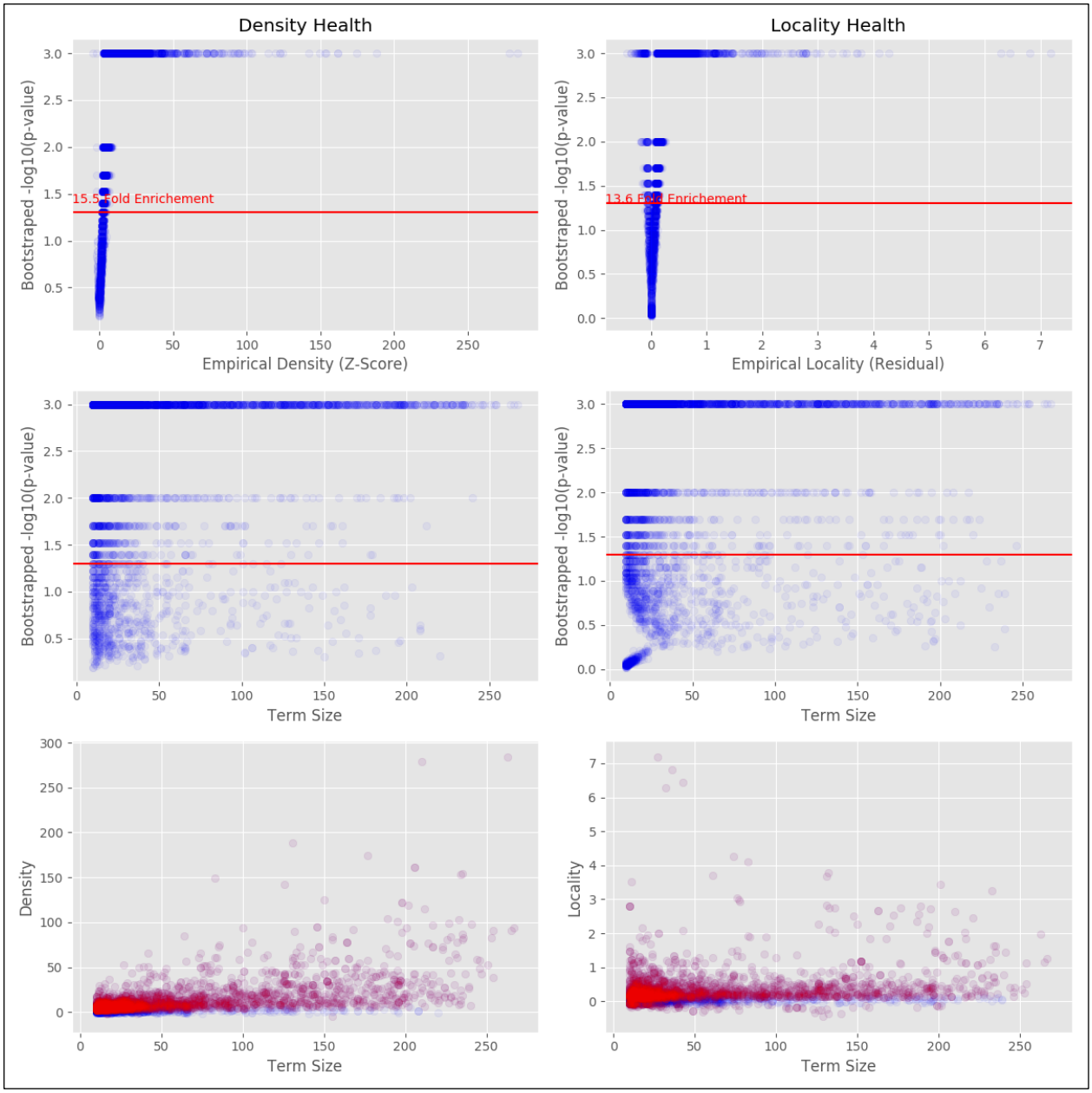


Fig S2. GO Term Co-Expression. The enrichment of co-expression for GO terms shown as a volcano plot, with three views for both of density and locality. The top two plots show expected vs. actual GO term co-expression. The bottom two rows of plots are a check to confirm there is no bias due to GO term size.

Table S1. Broad-sense Heritability (H^2) of Leaf Element Concentrations.

After outlier removal, lines with two or more replicates were used to calculate heritability for each trait.

Trait	H^2
SampleWeight	0.67
B11	0.61
Na23	0.81
Mg26	0.83
Al27	0.67
P31	0.85
S34	0.63
K39	0.83
Ca44	0.74
Fe54	0.74
Mn55	0.86
Co59	0.89
Ni60	0.81
Cu63	0.79
Zn66	0.76
As75	0.89
Se78	0.93
Rb85	0.82
Sr88	0.86
Mo98	0.91
Cd111	0.94

Table S2. QTL for Leaf Element Concentrations.

Significant QTL are listed with marker position, name, LOD score, permutation threshold from 1000 random permutations (a= 0.05), and effect direction. Traits that had no QTL are shown with "NA".

Trait Name	QTL.Name	Chr	Pos.cM	LOD Score	Marker Name	Perm Thresh	Effect Direction
Sample Weight	NA	NA	NA	NA	NA	NA	
B11	NA	NA	NA	NA	NA	NA	
Na23	4@196.9	4	196.90	4.12	SYN5595	3.64	B73
Mg26	NA	NA	NA	NA	NA	NA	
Al27	5@150.9	5	150.90	3.89	SYN24318	3.55	B73
P31	NA	NA	NA	NA	NA	NA	
S34	1@416.2	1	416.20	6.11	PZE-101218183	3.69	B73
K39	3@352.6	3	352.60	5.40	PZE-103163035	3.66	B73
Ca44	10@245.5	10	245.50	3.86	SYN19287	3.64	Mo17
Fe54	NA	NA	NA	NA	NA	NA	
Mn55	NA	NA	NA	NA	NA	NA	
Co59	1@396.9	1	396.90	3.70	SYN34771	3.58	B73
Ni60	9@7.7	9	7.70	6.08	PZE-109001536	3.69	B73
Cu63	3@163.7	3	163.70	4.11	SYN22860	3.66	B73
Zn66	1@401.0	1	401.00	5.26	SYN11	3.54	B73
As75	3@346.4	3	346.40	4.03	SYN32046	3.62	Mo17
Se78	NA	NA	NA	NA	NA	NA	
Rb85	NA	NA	NA	NA	NA	NA	
Sr88	NA	NA	NA	NA	NA	NA	
Mo98	1@378.0	1	378.00	16.71	SYN11473	3.64	Mo17
Cd111	1@406.0	1	406.00	5.07	SYN25458	3.59	B73
Cd111	2@214.6	2	214.60	21.06	SYN6540	3.59	Mo17

Table S3. Trans-eQTL Hotspots GO Enrichments.

Significantly enriched GO terms that pass multiple testing correction names and IDs are listed in the first two columns. Enrichment was calculated according to a hypergeometric calculation to measure the p-value of finding "num common" genes when you sample "source term size" and there are "target term size" genes of interest with "num universe" total genes. "Num terms" indicates the total number of terms in the gene ontology reference

name	id	pval	terms tested	num common	num universe	source term size	target term size	num terms	hotspot	multtest pass	genes	namespace
cohesin complex	GO:0008278	2.83E-06	367	4	39324	43	89	11909	HotspotChr2	TRUE	ZM00001D010854,ZM00001D036220, ZM00001D008541,ZM00001D050828	cellular_component
exodeoxyribonuclease activity	GO:0004529	6.37E-05	367	3	39324	34	89	11909	HotspotChr2	TRUE	ZM00001D036220,ZM00001D030663, ZM00001D050828	molecular_function
exodeoxyribonuclease activity, producing 5'-phosphomonoes	GO:0016895	6.37E-05	367	3	39324	34	89	11909	HotspotChr2	TRUE	ZM00001D036220,ZM00001D030663, ZM00001D050828	molecular_function
oxidoreductase activity, acting on the aldehyde or oxo group	GO:0016624	7.58E-05	367	3	39324	36	89	11909	HotspotChr2	TRUE	ZM00001D045745,ZM00001D014356, ZM00001D036104	molecular_function
Lys63-specific deubiquitinase activity	GO:0061578	1.04E-09	789	4	39324	5	151	11909	HotspotChr3	TRUE	ZM00001D052648,ZM00001D013752, ZM00001D040334,ZM00001D035432	molecular_function
proteasome regulatory particle, lid subcomplex	GO:0008541	2.92E-06	789	4	39324	26	151	11909	HotspotChr3	TRUE	ZM00001D052648,ZM00001D041456, ZM00001D040334,ZM00001D013752	cellular_component
phenylalanine ammonia-lyase activity	GO:0045548	1.01E-07	581	4	39324	19	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	molecular_function
cinnamic acid biosynthetic process	GO:0009800	1.90E-07	581	4	39324	22	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
cinnamic acid metabolic process	GO:0009803	2.30E-07	581	4	39324	23	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
ammonia-lyase activity	GO:0016841	4.52E-07	581	4	39324	27	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	molecular_function
erythrose 4-phosphate/phosphoenolpyruvate family amino a	GO:1902222	4.52E-07	581	4	39324	27	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
L-phenylalanine catabolic process	GO:0006559	4.52E-07	581	4	39324	27	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
L-phenylalanine metabolic process	GO:0006558	2.54E-06	581	4	39324	41	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
erythrose 4-phosphate/phosphoenolpyruvate family amino a	GO:1902221	2.54E-06	581	4	39324	41	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	biological_process
salicylic acid catabolic process	GO:0046244	2.60E-06	581	3	39324	12	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163	biological_process
phenol-containing compound catabolic process	GO:0019336	5.35E-06	581	3	39324	15	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163	biological_process
drought recovery	GO:0009819	1.79E-05	581	3	39324	22	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163	biological_process
carbon-nitrogen lyase activity	GO:0016840	1.95E-05	581	4	39324	68	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051166, ZM00001D051163,ZM00001D051164	molecular_function
coumarin biosynthetic process	GO:0009805	5.89E-05	581	4	39324	90	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051227, ZM00001D051166,ZM00001D051163	biological_process
coumarin metabolic process	GO:0009804	5.89E-05	581	4	39324	90	91	11909	HotspotChr4	TRUE	ZM00001D051161,ZM00001D051227, ZM00001D051166,ZM00001D051163	biological_process
histone acetyltransferase binding	GO:0035035	2.37E-10	406	4	39324	6	80	11909	HotspotChr5	TRUE	ZM00001D012941,ZM00001D002110, ZM00001D047596,ZM00001D034314	molecular_function
transcription factor TFIID complex	GO:0005669	1.52E-06	406	4	39324	41	80	11909	HotspotChr5	TRUE	ZM00001D047596,ZM00001D002110, ZM00001D034314,ZM00001D012941	cellular_component
guanosine-3',5'-bis(diphosphate) 3'-diphosphatase activity	GO:0008893	2.90E-06	406	3	39324	14	80	11909	HotspotChr5	TRUE	ZM00001D052381,ZM00001D005104, ZM00001D030043	molecular_function
diphosphoric monoester hydrolase activity	GO:0016794	2.90E-06	406	3	39324	14	80	11909	HotspotChr5	TRUE	ZM00001D052381,ZM00001D005104, ZM00001D030043	molecular_function
guanosine tetraphosphate metabolic process	GO:0015969	5.40E-06	406	3	39324	17	80	11909	HotspotChr5	TRUE	ZM00001D052381,ZM00001D005104, ZM00001D030043	biological_process
purine ribonucleoside bisphosphate metabolic process	GO:0034035	5.40E-06	406	3	39324	17	80	11909	HotspotChr5	TRUE	ZM00001D052381,ZM00001D005104, ZM00001D030043	biological_process
transcription factor binding	GO:0008134	1.03E-05	406	5	39324	140	80	11909	HotspotChr5	TRUE	ZM00001D002110,ZM00001D016456, ZM00001D012941,ZM00001D047596, ZM00001D034314	molecular_function
cortical cytoskeleton	GO:0030863	1.81E-05	406	3	39324	25	80	11909	HotspotChr5	TRUE	ZM00001D008768,ZM00001D038078, ZM00001D014053	cellular_component
protein O-linked fucosylation	GO:0036066	2.45E-05	406	2	39324	4	80	11909	HotspotChr5	TRUE	ZM00001D032064,ZM00001D015362	biological_process
transcription coactivator activity	GO:0003713	2.46E-05	406	4	39324	82	80	11909	HotspotChr5	TRUE	ZM00001D047596,ZM00001D002110, ZM00001D034314,ZM00001D012941	molecular_function
actin filament	GO:0005884	2.56E-05	406	3	39324	28	80	11909	HotspotChr5	TRUE	ZM00001D008768,ZM00001D038078, ZM00001D014053	cellular_component
actin filament depolymerization	GO:0030042	3.50E-05	406	3	39324	31	80	11909	HotspotChr5	TRUE	ZM00001D008768,ZM00001D038078, ZM00001D014053	biological_process
protein depolymerization	GO:0051261	4.23E-05	406	3	39324	33	80	11909	HotspotChr5	TRUE	ZM00001D008768,ZM00001D038078, ZM00001D014053	biological_process
actin filament binding	GO:0051015	4.39E-05	406	4	39324	95	80	11909	HotspotChr5	TRUE	ZM00001D008768,ZM00001D038078, ZM00001D014376,ZM00001D014053	molecular_function
RNA polymerase III transcription factor complex	GO:0090575	0.00011995	406	4	39324	123	80	11909	HotspotChr5	TRUE	ZM00001D047596,ZM00001D002110, ZM00001D034314,ZM00001D012941	cellular_component
DNA packaging	GO:0006323	3.55E-08	498	5	39324	37	98	11909	HotspotChr6	TRUE	ZM00001D020387,ZM00001D010608, ZM00001D050017,ZM00001D017576, ZM00001D039978	biological_process
acireductone dioxygenase [iron(II)-requiring] activity	GO:0010309	7.27E-07	555	3	39324	9	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	molecular_function
heteropolysaccharide binding	GO:0010297	1.04E-06	555	3	39324	10	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	molecular_function
L-methionine biosynthetic process from methylthioadenosine	GO:0019509	4.80E-06	555	3	39324	16	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	biological_process
amino acid salvage	GO:0043102	5.82E-06	555	3	39324	17	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	biological_process
L-methionine salvage	GO:0071267	5.82E-06	555	3	39324	17	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	biological_process
oxidoreductase activity, acting on single donors with incorpor	GO:0016702	3.23E-05	555	5	39324	173	82	11909	HotspotChr9	TRUE	ZM00001D033377,ZM00001D041103, ZM00001D004756,ZM00001D009286, ZM00001D019074	molecular_function
L-methionine biosynthetic process	GO:0071265	5.45E-05	555	3	39324	35	82	11909	HotspotChr9	TRUE	ZM00001D019074,ZM00001D004756, ZM00001D041103	biological_process
oxidoreductase activity, acting on single donors with incorpor	GO:0016701	7.91E-05	555	5	39324	209	82	11909	HotspotChr9	TRUE	ZM00001D033377,ZM00001D041103, ZM00001D004756,ZM00001D009286, ZM00001D019074	molecular_function
SCF complex assembly	GO:0010265	8.96E-05	555	2	39324	7	82	11909	HotspotChr9	TRUE	ZM00001D049167,ZM00001D027974	biological_process
shikimate 3-dehydrogenase (NADP+) activity	GO:0004764	4.75E-07	571	3	39324	7	95	11909	HotspotChr10	TRUE	ZM00001D023895,ZM00001D023892, ZM00001D023888	molecular_function
3-dehydroquinate dehydratase activity	GO:0003855	4.75E-07	571	3	39324	7	95	11909	HotspotChr10	TRUE	ZM00001D023895,ZM00001D023892, ZM00001D023888	molecular_function
riboflavin synthase activity	GO:0004746	5.77E-06	571	2	39324	2	95	11909	HotspotChr10	TRUE	ZM00001D023836,ZM00001D023863	molecular_function
cell motility	GO:0048870	2.90E-05	571	4	39324	72	95	11909	HotspotChr10	TRUE	ZM00001D023905,ZM00001D031846, ZM00001D023906,ZM00001D023903	biological_process

Table S4. Trans-eQTL Hotspots MCL Cluster Enrichments.

Significantly enriched MCL clusters that pass multiple testing correction are shown in the first column. Enrichment was calculated using a hypergeometric calculation to measure the p-value of finding "num common" genes when you sample "source term size" and there are "target term size" genes of interest with "num universe" total genes. "Num terms" indicates the total number of terms in all MCL clusters.

id	pval	terms tested	num common	num universe	source term size	target term size	num terms	hotspot	multi test pass
MCL28	1.34E-78	9	36	24430	53	89	4042	HotspotChr2	TRUE
MCL70	4.59E-34	9	16	24430	25	89	4042	HotspotChr2	TRUE
MCL85	5.39E-22	9	11	24430	22	89	4042	HotspotChr2	TRUE
MCL76	2.20E-12	9	7	24430	24	89	4042	HotspotChr2	TRUE
MCL27	7.12E-90	22	44	24430	55	151	4042	HotspotChr3	TRUE
MCL40	3.33E-63	22	32	24430	43	151	4042	HotspotChr3	TRUE
MCL48	3.88E-42	22	23	24430	38	151	4042	HotspotChr3	TRUE
MCL155	8.92E-25	22	12	24430	15	151	4042	HotspotChr3	TRUE
MCL565	2.26E-04	22	2	24430	4	151	4042	HotspotChr3	TRUE
MCL344	7.81E-04	22	2	24430	7	151	4042	HotspotChr3	TRUE
MCL60	1.32E-51	10	23	24430	30	91	4042	HotspotChr4	TRUE
MCL49	1.95E-40	10	20	24430	36	91	4042	HotspotChr4	TRUE
MCL87	4.39E-17	10	9	24430	22	91	4042	HotspotChr4	TRUE
MCL79	2.58E-12	10	7	24430	24	91	4042	HotspotChr4	TRUE
MCL196	4.98E-10	10	5	24430	12	91	4042	HotspotChr4	TRUE
MCL467	8.98E-10	10	4	24430	5	91	4042	HotspotChr4	TRUE
MCL506	4.97E-07	10	3	24430	5	91	4042	HotspotChr4	TRUE
MCL43	1.82E-48	11	23	24430	40	80	4042	HotspotChr5	TRUE
MCL59	1.07E-35	11	17	24430	30	80	4042	HotspotChr5	TRUE
MCL86	4.04E-25	11	12	24430	22	80	4042	HotspotChr5	TRUE
MCL92	8.36E-09	11	5	24430	22	80	4042	HotspotChr5	TRUE
MCL423	6.71E-07	11	3	24430	6	80	4042	HotspotChr5	TRUE
MCL149	1.08E-03	11	2	24430	15	80	4042	HotspotChr5	TRUE
MCL30	2.68E-102	9	44	24430	52	98	4042	HotspotChr6	TRUE
MCL32	8.29E-46	9	24	24430	49	98	4042	HotspotChr6	TRUE
MCL83	8.79E-17	9	9	24430	22	98	4042	HotspotChr6	TRUE
MCL639	9.51E-05	9	2	24430	4	98	4042	HotspotChr6	TRUE
MCL379	3.30E-04	9	2	24430	7	98	4042	HotspotChr6	TRUE
MCL24	2.70E-56	15	28	24430	59	78	4042	HotspotChr7	TRUE
MCL176	1.51E-20	15	9	24430	13	78	4042	HotspotChr7	TRUE
MCL66	1.58E-14	15	8	24430	27	78	4042	HotspotChr7	TRUE
MCL431	1.74E-12	15	5	24430	6	78	4042	HotspotChr7	TRUE
MCL156	1.11E-05	15	3	24430	14	78	4042	HotspotChr7	TRUE
MCL38	7.21E-52	16	25	24430	45	82	4042	HotspotChr9	TRUE
MCL81	9.02E-45	16	19	24430	23	82	4042	HotspotChr9	TRUE
MCL283	4.69E-11	16	5	24430	9	82	4042	HotspotChr9	TRUE
MCL640	1.45E-07	16	3	24430	4	82	4042	HotspotChr9	TRUE
MCL129	1.46E-03	16	2	24430	17	82	4042	HotspotChr9	TRUE
MCL25	2.73E-70	7	34	24430	56	95	4042	HotspotChr10	TRUE
MCL34	1.31E-54	7	27	24430	47	95	4042	HotspotChr10	TRUE
MCL37	1.81E-21	7	13	24430	46	95	4042	HotspotChr10	TRUE
MCL205	8.04E-04	7	2	24430	11	95	4042	HotspotChr10	TRUE
MCL61	6.06E-03	7	2	24430	30	95	4042	HotspotChr10	TRUE

CHAPTER 5:
CONCLUSIONS AND FUTURE DIRECTIONS

In Chapter 2, I demonstrated that the maize kernel ionome is determined by genetic and environmental factors, with a large number of genetic by environment interactions. Elemental profiling of the IBM population across 10 environments allowed us to capture environmentally-driven variation in the ionome. The QTL analysis on elements found mainly single-environment QTL, indicative of substantial genetic by environment interaction in establishment of the elemental composition of the maize grain. This approach, along with identification of QEI occurring both within a single location over different years and QEI between different locations, indicated that gene by environment interactions underlie elemental accumulation in maize kernels.

In Chapter 3, I expanded the element QTL analysis to include variables representing the network properties of the ionome. Using this approach showed that treating the ionome as an interrelated set of traits through PCA within environments can identify novel loci. PCA across environments allowed us to derive traits that described both environmental and genetic variation in the ionome. While the multiple environment analyses here were limited by the lack of environmental data collected during the growing season, future experiments could apply the same multivariate technique to distinguish environments based on the whole ionome and test to see which environmental variables are driving contrasts. Studies across a larger set of environments, with soil and weather data measured consistently throughout the growing season, can use this multivariate approach as well as include specific environmental variables in QTL models to model QTL interactions with particular environmental components.

In Chapter 4, I incorporated gene expression data collected from roots of the IBM population. The roots are a key tissue, if not the primary tissue, in shaping the ionome of the whole plant. Specifically, gene expression changes in the root have been shown to alter the ionome. Using eQTL mapping, I found associations between gene expression variation in the root and variation at genetic loci, some of which were loci previously linked to the ionome. A locus of special interest is the cadmium QTL on chromosome 2, a region that is also a trans-eQTL hotspot with genes correlated with Cd among its set of gene targets. The gene expression study supplied additional support for investigating the causal gene or genes under this QTL. Our group is currently working on fine-mapping this locus by developing near-isogenic lines (NILs) which break up recombination in the region of interest in a consistent genetic background. In these NILs, we can profile Cd accumulation and perform RNA-seq and eQTL mapping, essentially the same process as conducted before but with a more defined region and higher genetic resolution. Approaches such as this that refine genetic regions and test genes can be utilized for other QTL and candidate gene lists.

This thesis has set forth an integrative approach to understanding element accumulation in maize. The genetic basis of complex traits is challenging to dissect and requires a combination of multiple -omics and phenotyping approaches. QTL mapping in the maize seed and leaf, followed by transcriptome-based analysis in the root, where gene expression changes often influence the seed and leaf ionomes, identified a set of candidate genes for regulation of elements that can be further explored to improve models of element homeostasis and develop agricultural applications.