bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The Interaction of Genotype and Environment Determines Variation in the Maize Kernel Ionome

Alexandra Asaro[1], Greg Ziegler[2], Cathrine Ziyomo[1], Owen A. Hoekenga[3], Brian P. Dilkes[4], Ivan Baxter[2]*

[1]Donald Danforth Plant Science Center, St. Louis, Missouri, United States of America

[2]USDA-ARS, Donald Danforth Plant Science Center, St. Louis, Missouri, United States of America

[3]Genomics Consultant, Ithaca, New York, United States of America

[4]Department of Biochemistry, Purdue University, West Lafayette, Indiana, United States of America

*Corresponding author

E-mail: ivan.baxter@ars.usda.gov (IB)

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Plants obtain soil-resident elements that support growth and metabolism via water-mediated flow facilitated by transpiration and active transport processes. The availability of elements in the environment interact 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 influencing elemental accumulation 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. We identified quantitative trait loci controlling elemental accumulation by considering individual elemental accumulation measurements as traits and by mapping the loci responsible for variation in co-regulated multi-elemental traits identified using principle components analysis. These approaches detected partially overlapping sets of loci, many of which were found only in a single growout. We applied our multi-element approach across all of the growouts and found that the growth environment has a profound effect on the elemental profile and that some multi-element phenotypes correlate with specific environmental variables. We identified QTL by environment interactions (QEIs) through three methods: linear modeling with environmental covariates, QTL analysis on trait differences between growouts, and QTL analysis on factors obtained from a principle component derived model of ionome variation across environments. Overall, we were able to map 79 elemental QTL, 101 principal component QTL, and several instances of QEI, indicating that elemental profiles are highly heritable, interrelated, and responsive to the environment.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Summary

Plants take up elements from the soil, a process that is highly regulated by the plant's genome. In order to look at how maize alters its elemental uptake in response to different environments, we analyzed the kernel elemental content of a population derived from a cross grown 10 different times in six locations. We found that environment had a profound effect on which genetic loci were important for elemental accumulation in the kernel. We also found that the elements are not regulated independently and that mathematical combinations of elements will identify different genetic loci than single element approaches. The mathematical combinations of elements are correlated with environmental variables, suggesting that underlying the observed variation are interactions between genetically controlled factors and environmental variables. Our results suggest that to have a full understanding of elemental accumulation in maize kernels and other food crops, we will need to understand the interactions identified here at the level of the genes and the environmental variables that contribute to loading essential nutrients into seeds.

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 both to the availability of each regulated element in the environment and to the levels of these elements at the sites of use within the plant. Understanding how the genome encodes responses to element

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

limitation or toxic excess in nutrient-poor or contaminated soils will help 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.

There are many lines of evidence that elements covary with each other due to physiological, genetic or environmental factors. For example, a suite of elements responds to Fe deficiency in such a concerted manner that they can be used to predict the Fe status of the plant [4]. When *A. thaliana* recombinant populations were grown in multiple environments, genetic correlations among Li-Na, Mg-Ca, and Cu-Zn were observed across all environments while other pairs (Ca-Fe and Mg-Fe) were only correlated in a subset of environments [5]. These data indicate that, while understanding the factors driving individual element accumulation is important, it is also necessary to consider the ionome as a network of co-regulated and interacting traits [6]. Multivariate analysis techniques, such as principal components analysis (PCA), reduce multivariate data dimension by minimizing the variances of multiple input factors to new variables. When multiple phenotypes are correlated, such as for multiple elements in the ionome, this approach may prove to be complementary to single element approaches so as to better summarize these relationships.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [7]. 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 analyzed using both single and multivariate approaches to identify QTL and QEI underlying elemental accumulation. By seeking both environmental and genetic determinants, we detected loci controlling elemental accumulation, many of which were environment-specific. We also show that multivariate approaches can reveal environmental and genetic effects that cannot be detected using single element approaches.

Results

Genetic Regulation of Elemental Traits

The data used for this study is comprised of 20 elements measured in the seeds from *Zea mays* L. Intermated B73 x Mo17 recombinant inbred line (IBM) populations 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 [8]. The four rounds of intermating and subsequent inbreeding resulted in more 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 [9-14]. This greater resolution reduces the number of genes

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

within a QTL support interval and increases the utility of QTL mapping as a hypothesis test for shared genetic regulation of multiple traits and aids in the 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 Ukilima, South Africa in 2010 (SA10) (Table S1). Single seeds were profiled for the quantities of 20 elements using ICP-MS and these measurements were normalized to seed weight and technical sources of variation using a linear model [15]. These normalized values are referred to as the elemental traits.

Variation in the elemental traits was affected by both environment and genotype. All elemental traits exhibited greater variation among genotype replicates grown across multiple environments than among replicates within a single environment (Table 1). The broad-sense heritability of seed weight, 15 of 21 elements in NY05, 13 of 21 elements in NC06, and 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 of element concentrations.

Trait	All env	NY05	NC06	MO06

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Seed Weight	0.51	0.59	0.69	0.89
B	0.02	0.35	0.51	0.06
Na	0.11	0.34	0.23	0.19
Mg	0.04	0.77	0.69	0.75
Al	0.10	0.39	0.50	0.08
P	0.04	0.62	0.69	0.33
S	0.05	0.73	0.77	0.51
K	0.07	0.69	0.72	0.36
Ca	0.15	0.65	0.63	0.77
Mn	0.16	0.80	0.80	0.75
Fe	0.07	0.76	0.73	0.63
Co	0.08	0.65	0.54	0.42
Ni	0.06	0.84	0.54	0.82
Cu	0.20	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	0.04	0.32	0.35	0.68
Rb	0.03	0.49	0.45	0.69
Sr	0.07	0.61	0.48	0.53
Mo	0.29	0.85	0.73	0.96
Cd	0.55	0.71	0.69	0.24

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 [16], 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 for each marker on a phenotype. Forward and backward regression was used to generate final genome-wide QTL models for each trait. This QTL mapping procedure was completed for each of the IBM populations from each of the 10 environments for all 21 traits as separate analyses. QTL significance were determined using a penalty score for adding QTL to the stepwise model derived by taking the 95th percentile of the highest LOD score achieved in 1000 runs of the *scanone* genome scan with random data [17].

The environmental dependence on QTL detection was first estimated by identifying QTL

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

common to multiple environments. If QTL detected in two or more growouts affected the same element and localized less than 5 cM apart they were considered to be the same locus. A total of 79 QTL were identified for seed weight and the 20 elemental traits across these 10 environments (Fig 1B &C). Of these, 63 were detected in a single environment and 16 QTL 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 [4], 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).

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

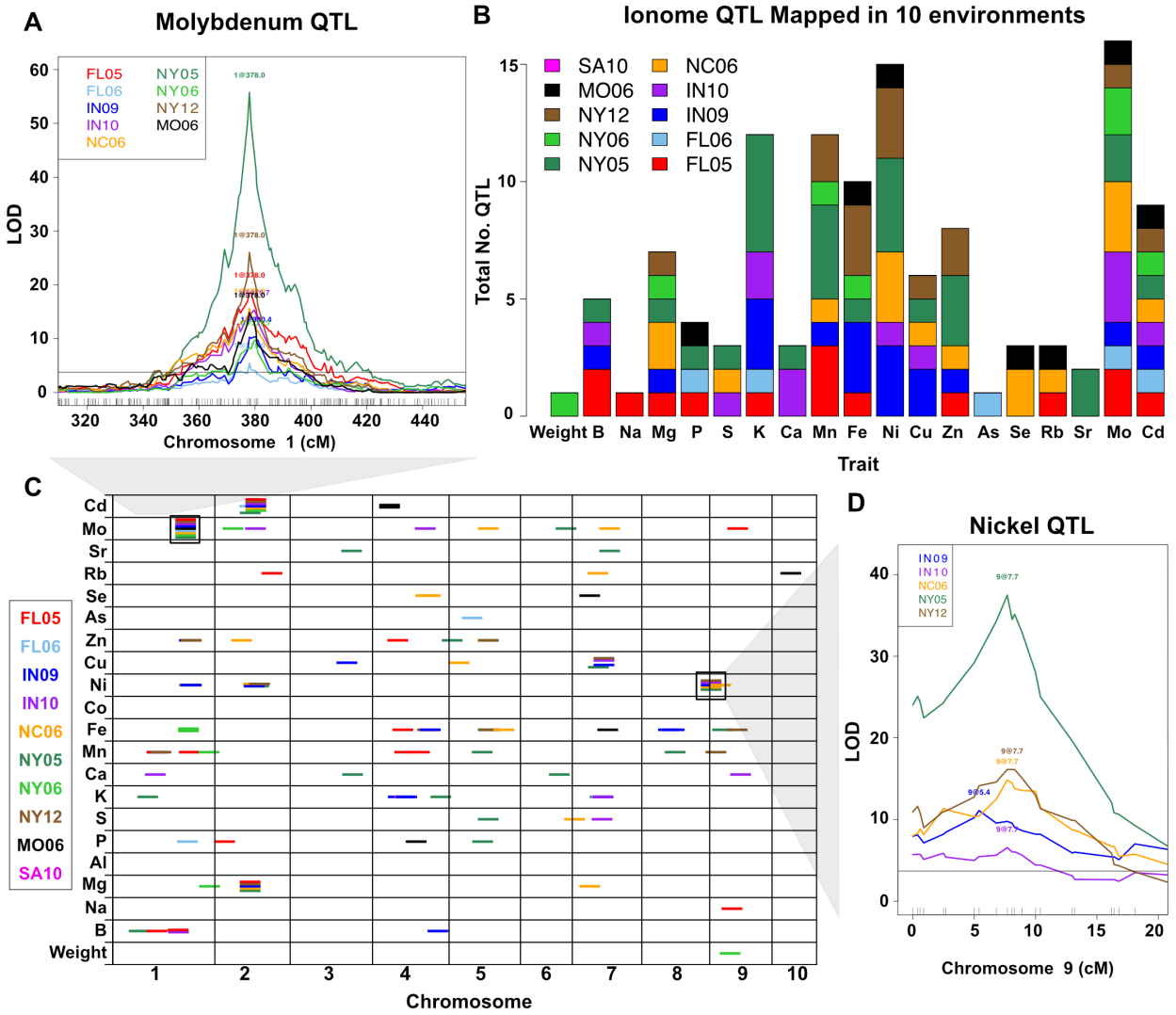


Fig 1. Ionome QTL from 10 Environments. QTL identified for seed weight and 20 element accumulation traits using the B73 x M017 intermated RIL population grown in 10 environments. (A) QTL on chromosome 1 affecting variation in molybdenum accumulation. An interval of Chr1 is shown on the x-axis (in centi-Morgans). The LOD score for the trait-genotype association is shown on the y-axis. The horizontal line is a significance threshold corresponding to the 95th percentile of highest LOD score from 1000 random permutations. 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 (loci with LOD scores at or above the 95th percentile of the highest LOD score achieved by running the stepwise QTL mapping algorithm on 1000 random permutation replicates) 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.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 negatives assessment of QTL absence due to genetic action but statistical assessment below the permutation threshold in other locations. 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 location-specific QTL remained significant. Using a Bonferroni correction for 200 tests (20 traits in 10 environments), 10 QTL (95th percentile threshold) and two QTL (99th percentile threshold) should exceed this threshold. The presence of 63 and 31 QTL greatly exceeds these null approximations. 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.

Accounting for Element to Element Correlations

Previous elemental correlation studies and mutant phenotype analyses indicate extensive relationships between elements [4,6]. In this experiment, several elements were highly correlated across the dataset and nine loci were found with QTL for two or more different elements (Table 2). Phosphorous exhibited the highest incidence of shared QTL with other elements, including shared QTL with the cations K and Mg and the only shared QTL found between more than two elements, between P, S, Fe, Mn, and Zn in NY05 (Fig 2). Shared QTL were also found between elements with similar structure, charge, and/or type, such as Ca and Sr or Fe and Zn.

Table 2. QTL affecting variation for Multiple Elements in the same environment.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

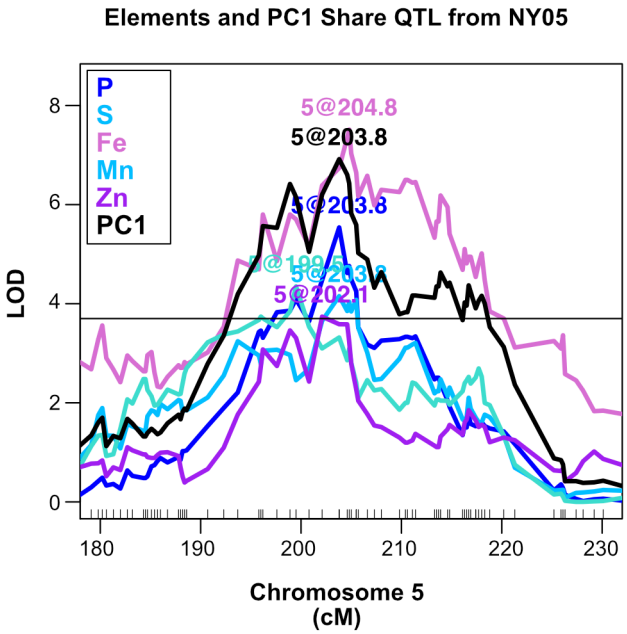


Fig 2. Multiple Element QTL. Stepwise QTL mapping output from the NY05 population for P, S, Fe, Mn, Zn, and PC1. Position in cM on chromosome 5 is plotted on the x-axis and LOD score is shown on the y-axis. 95th percentile of highest LOD score from 1000 random permutations is indicated as horizontal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

To better identify factors underlying multi-element correlation, principal components analysis (PCA) was used. This reduced correlated elements into principal components (PCs), orthogonal variables that account for variation in the original dataset, each having an associated set of rotations from the input variables. PCA was conducted in each of the 10 environments separately. Elements that were difficult to measure and potentially introduce artifacts (B, Na, Al,

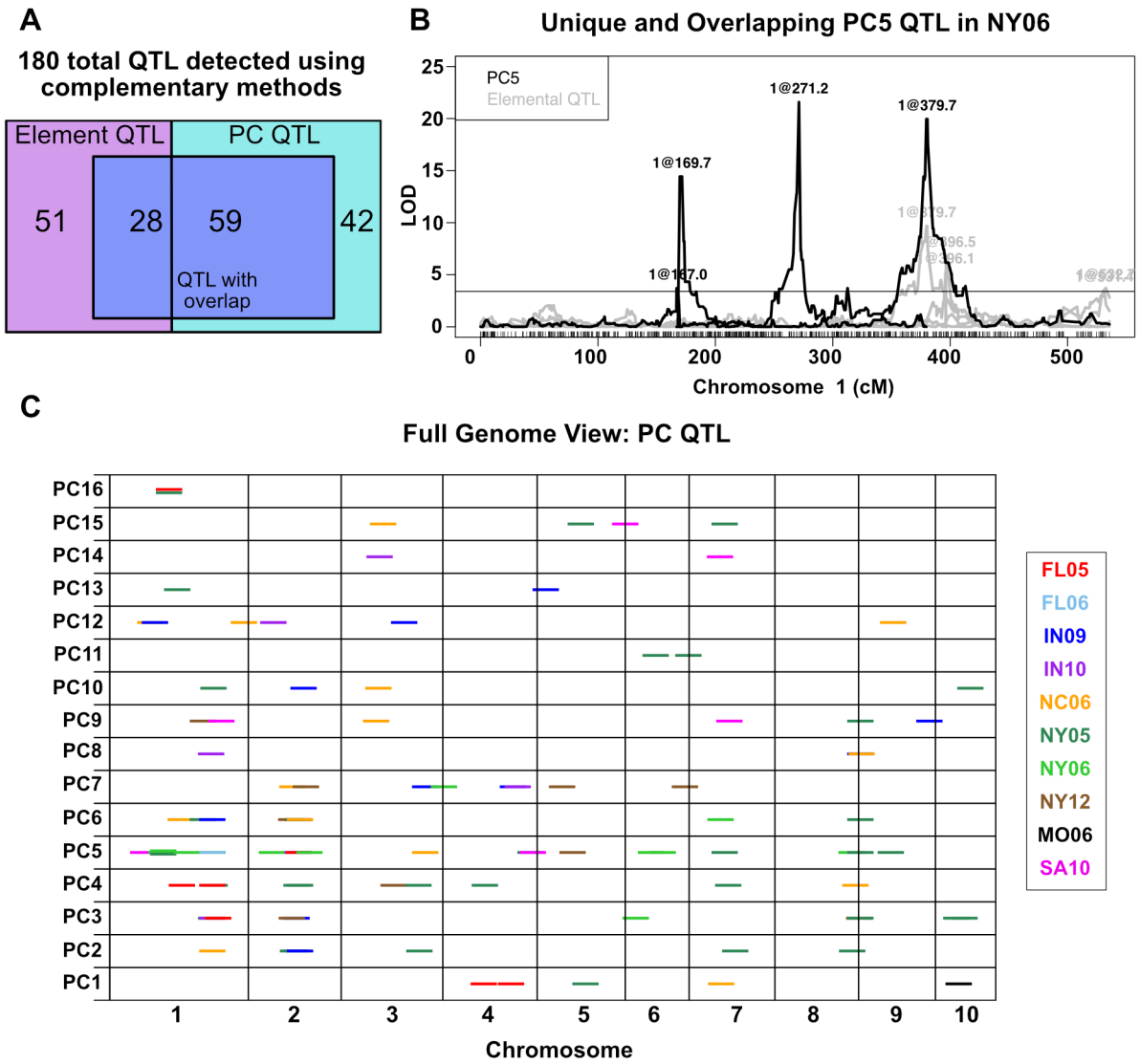
bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As) were excluded, as their covariance due to shared technical variation is problematic. The remaining elemental data from lines within a single environment were transformed into 16 principal components (Fig S1). These 10 sets of 16 PCs were then used as traits for QTL mapping.

QTL mapping using these derived traits yielded 101 QTL (Fig 3C). PC QTL overlapped with many single element QTL, consistently recapitulating strong single element QTL seen for elements such as Ni, Mo, and Cd. The PC QTL also capture previously observed multi-element QTL. For example, in NY05, a QTL for PC1 overlaps the QTL that were detected in the single element analyses of P, S, Fe, Mn, and Zn on chromosome 5 (Fig 2). The PC QTL in this case is as strong as the Fe QTL and much more significant than the P, S, Mn, and Zn elemental QTL, suggesting that a PC capturing multi-element variation may strengthen the signal acquired using a single-element approach. In addition, mapping with PCs allowed for identification of new loci not seen using single elements. QTL mapping on single elements may not have the power to detect loci with small effects on several elements, so the unique QTL detected using PCs as traits are of particular interest. For instance, two PC5 QTL from the NY06 growout on chromosome 1 were distinct from any elemental QTL (Fig 3B). 42 PC-specific QTL, defined as QTL >25 cM away from any elemental QTL in the same environment, were detected. PCs are derived from combinations of elements, and as a result single elemental QTL were reproduced for several PCs within an environment. This is observed particularly for elements with strong single-element effects, such as Mo and Cd. For example, in IN10, PC2 and PC10 have QTL that co-localize with the same Cd QTL on chromosome 2. Likewise, in NY05, PC3, PC5, PC6, and PC9 all have QTL that overlap with a single chromosome 9 Ni QTL. This redundancy contributes to the higher number and proportion of detected PC QTL shared with element QTL (59/101) than

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

220 element QTL shared with PC QTL (28/79), although the same genomic locations underlie this
221 overlap (Fig 3A). The identification of unique QTL through this multivariate approach
222 demonstrates the complementary nature of working with trait covariance as well as the
223 component traits and shows that elemental traits are mechanistically interrelated [6,18].



224

225 **Fig 3. Principal Component QTL from 10 environments.** PCs were derived from elemental
226 data from the IBM mapping population lines separately in each of 10 environments. 10 sets of 16
227 PCs, each set from a single environment, were used as traits for QTL mapping. (A) 180 total
228 element and PC QTL were mapped. The two boxes represent the 79 and 101 elemental and PC
229 QTL, respectively. 28 element QTL overlap with PC QTL from the same environment. 59 PC
230 QTL overlap with element QTL from the same environment. These sets of non-unique QTL are
231 shown in the box spanning both boxes. QTL unique to elements, 51, and to PCs, 42, are shown

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

outside of the shared box. (B) QTL mapping output for PC5 from the NY06 population. Chromosome 1 is shown on the x-axis (in cM), LOD score is on the y-axis. Significance threshold is calculated as 95th percentile of highest LOD score from 1000 random permutations. All significant NY06 element QTL on chromosome 1 are shown in grey. 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 (loci with LOD scores at or above the 95th percentile of the highest LOD score achieved by running the stepwise QTL mapping algorithm on 1000 random permutation replicates) for 16 PCs in 10 environments. QTL location is shown across the 10 maize chromosomes (in cM) on the x-axis. Environment in which QTL was found is designated by color. QTL are represented as dashes of uniform size.

QTL by Environment Interactions

That QTL detection was so strongly affected by environment suggested that the effects of allelic variation were heavily dependent on environments for both elemental and PCA traits. These results, however, did not specifically test for QTL by environment interactions (QEI). Comparison between environments in our data is additionally complicated because different subsamples of the IBM population were grown at these different locations and years. There are many different approaches to identifying QEI described in the literature, summarized in El-Soda et al. We took two previously implemented methods of QEI analysis. 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 genetically identical RIL 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]. We added a third approach, not previously described, to determine the gene by environment effects on the multi-element phenotypes. We extended the PCA analysis across years and locations by calculating PCs on a subset of lines across environments, and using the loadings to calculate a projection for each line in each environment. If the genetic and environmental variances do not interact, some projections will reflect environmental variance

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 projections calculated from ionomic traits should be both correlated with environmental factors and result in detectable QTL.

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), averaging common lines across years.

Two linear equations were fit to the combined data using the FL, IN, and NY locations as covariates. The first equation is the full model considering phenotypic variation as controlled by genotype and location as both additive and interacting covariates (1) while the reduced models consider phenotypic variation as resulting from location and genotype as additive factors (2). Subtracting (2) from (1) isolates genetic by location variation.

$$V_P = V_G + V_E + V_{GxE} \qquad (1)$$

$$V_P = V_G + V_E \qquad (2)$$

$$V_P = V_{GxE} \qquad (1) - (2)$$

The program R/qtl was used to fit QTL for phenotypic variation with and without considering QTL by location interactions for sample weight and 20 elements. The significance threshold was calculated by using the 95th percentile of the highest LOD score from 1000

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

permutations of the three step procedure (fitting the two models and then subtracting LOD scores). Even with this underpowered dataset, 10 QTL by location interactions exceeded this threshold (Table 3). 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 (Fig 4). These QTL corresponded to loci with significant QTL by location effect (Fig 4). 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. This 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 3. QTL with Significant by-Location interactions.

Trait	Chr	Pos (cM)	LOD	Significance Threshold†
Na	3	237.00	3.48	3.39
Na	9	115.40	4.24	3.39
Mn	1	232.00	4.77	3.84
Fe	5	196.20	3.69	3.57
Ni	9	7.70	23.59	5.44
Zn	4	157.40	3.82	3.78
Rb	2	188.40	4.30	3.41
Rb	10	93.60	4.25	3.41

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mo	1	378.00	32.85	4.32
Cd	2	214.60	14.47	3.59

†95th percentile LOD score from 1000 random permutations

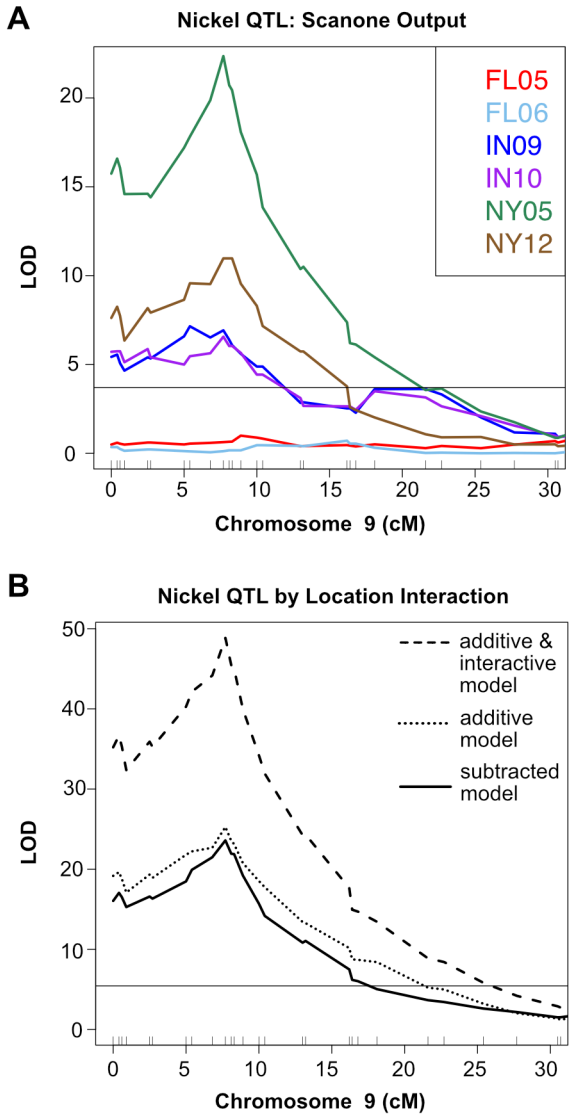


Fig 4. Significant QTL-by-Location Interactions Reflect Variation in Single Environment Mapping. (A) Nickel QTL on chromosome 9 exhibits variation between FL, IN, and NY growouts in single environment QTL mapping. Scanone QTL mapping output for Ni on is plotted for FL05, FL06, IN09, IN10, NY05, and NY12. LOD score is plotted on the y-axis and cM position on the x-axis. Horizontal line corresponds to 95th percentile of highest LOD score from 1000 random permutations. (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 a location were included, with values averaged between lines common to multiple years in a 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 shown as dashed line. Subtracted LOD score profile from the two models

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(QTL by location interactive effect only) 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 the intra-location QEI in the three locations (FL05 & FL06, IN09 & IN10, NY05 & NY12). Using the same stepwise algorithm that was implemented with single element and PC analysis, QTL were mapped on the trait differences between common lines in the two environments for sample weight and 20 elements. Mapping the trait differences between years for the three locations 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 4). These trait-difference QTL included locations identified in our single element and single 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 5A, B, C). Six QTL were detected that affected variation in the year to year difference but no QTL were detected at that location when the years were mapped separately. This demonstrates that this method can detect by-year differences that were not apparent by contrasting QTL detected from each year's 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 4. Significant QTL for Trait Differences.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

†95th percentile LOD score from 1000 random permutations

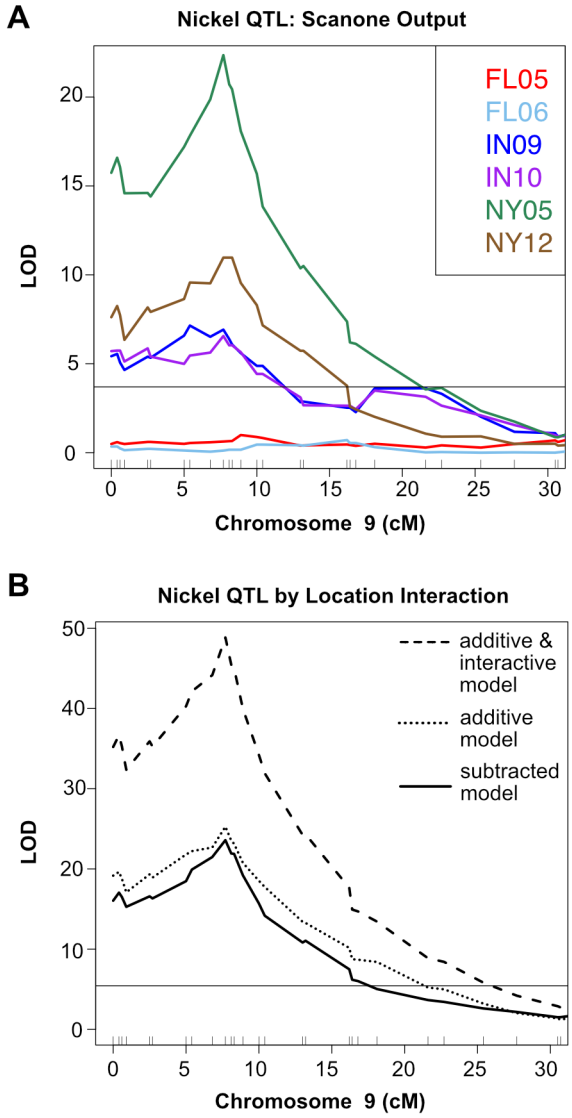


Fig 5. Comparison of QTL Mapped on Traits in Single Environments and Trait Differences Between Environments. Examples from stepwise QTL mapping on trait differences of between two years at one location, calculated between IBM lines common to both

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

years. Scanone QTL mapping output is plotted for the same trait from each year separately. LOD score is plotted on the y-axis and cM position on the x-axis. Horizontal lines correspond to 95th percentile of highest LOD score from 1000 random permutations. (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.

PCA-derived projections. 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. Only the 16 lines common to six of the 10 environments (FL05, FL06, IN09, IN10, NY05, NY12) were used to calculate the PCA. The rotations from this PCA were then projected onto the centered and scaled full dataset. In this way, 16 projections (PRs) derived from PCs describing the covariation of the ionome were calculated for every RIL in every environment. To reduce the incidence of artifacts or over fitting, PRs accounting for less than 2% of the total variation were eliminated, leaving seven PRs (Fig S2). The covariation captured by these PRs could due to genetic variation, result from a programmed response of the plant across multiple environments, or a combination of both.

The first two PRs were highly responsive to the environment (Fig 6). The lines from each environment cluster together when plotting PR1 vs PR2 values, with distinct separation between environments and years. Additionally, when modeled as a simple linear function, environment had a significant effect on all PRs ($p<0.001$). In order to identify environmental factors responsible for ionome covariance, weather station data from each location and year was collected from Climate Data Online (CDO) (http://www.ncdc.noaa.gov/cdo-web/). Average minimum temperature, maximum temperature, and growing degree days (GDD = ((Tmax + Tmin)/2) –10) across the entire 120-day growing season and over each 30-day quarter were calculated from daily summaries (Fig. 7A). Correlations were calculated between weather

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variables and average PR 1-7 values in the nine environments with weather data. Correlations were detected between PRs and weather variables. The weather variables, all temperature-based, are not correlated with PRs in many cases, although correlations exceeding r_p= 0.60 were observed for PRs 4-7 (Fig. 7A, Table S3). The strongest correlation observed for PR1 was with average maximum temperature in the fourth quarter of the growing season (r_p = 0.37) (Fig 7B) while the highest observed for PR2 was for average minimum temperature during the fourth quarter (r_p = 0.47) (Fig 7C). The relatively small number of environments, interrelatedness of the weather variables, and likely contribution of factors other than temperature limit the descriptive power of these correlations. The environmental components of projection variables are unknown but, consistent with element remobilization from the leaves to the seeds, temperature during the fourth quarter of the growing season may be one contributor to the environmental covariance in the seed ionome. The lack of particularly strong correlations between the first two PRs and temperature variables suggests that non-recorded variables, such as field to field variation in soil composition, fertilizer application, humidity, or biotic factors are likely to have an impact.

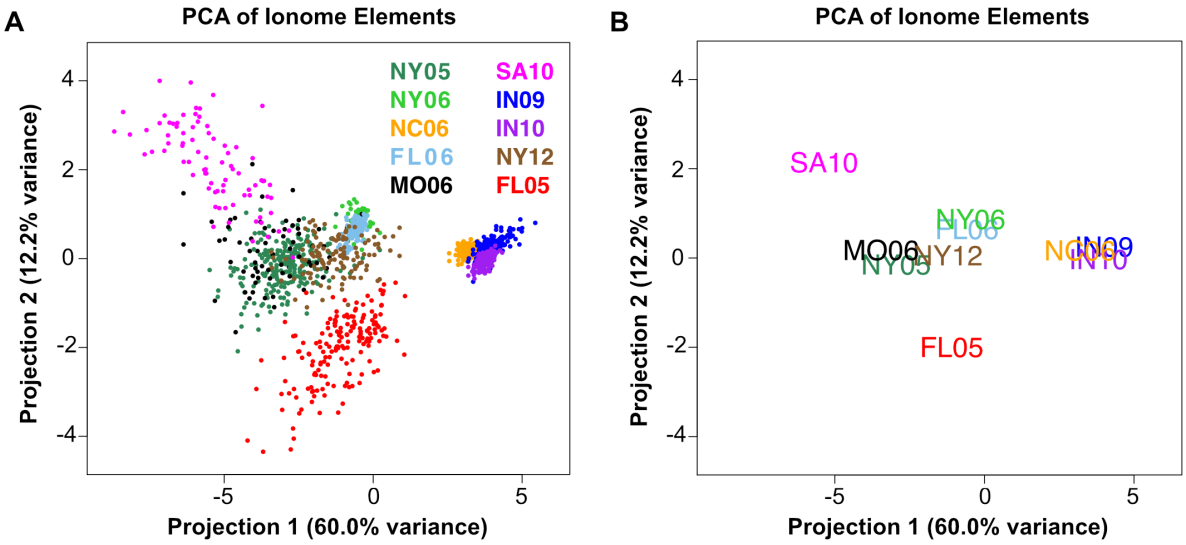


Fig 6. PC-Based Projections Separate Lines by Environment. PR1 and PR2 separate lines by environment. Points correspond to lines, colored by their environment. (A) PR1 vs PR2 values

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for each line, colored by environment. Percentage of total variance accounted for by each PR indicated on the axes. (B) Average PR1 vs PR2 values for all lines in each environment.

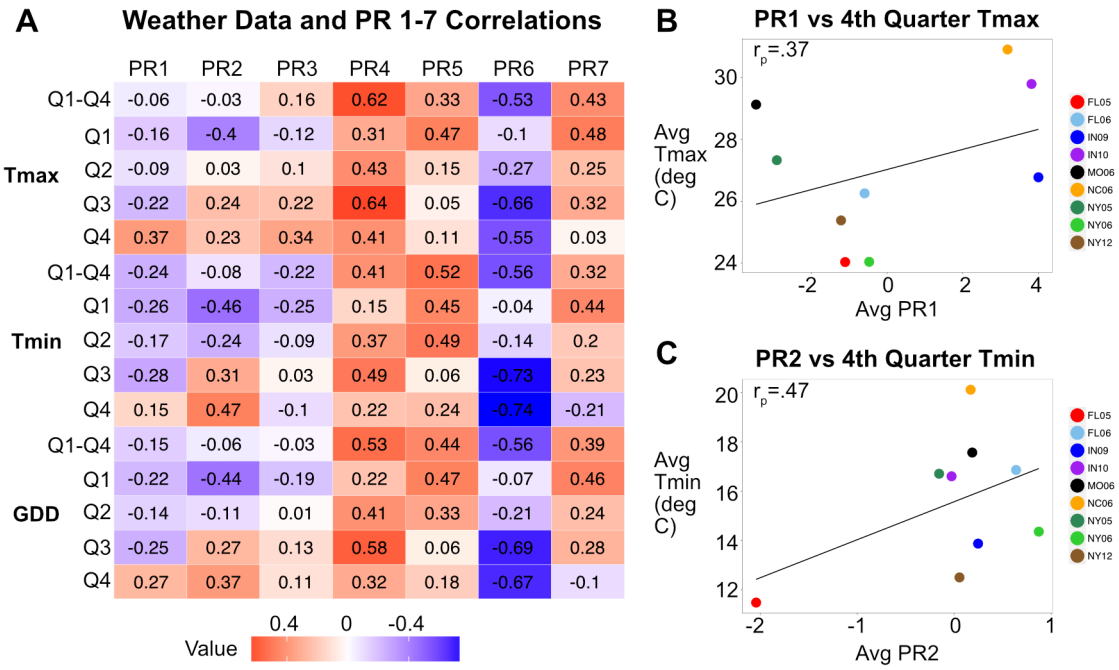


Fig 7. PR and Weather Variable Correlations. (A) Heatmap showing correlations between averaged PR 1-7 values across environments and averages for maximum temperature, minimum temperature, and growing degree days (GDD = (Tmax + Tmin)/2 –10)) across the entire growth season and for each quarter of the growth season. Red intensity corresponds to size of positive correlations. Blue intensity corresponds to magnitude of negative correlations. (B) Averaged PR1 values for 9 environments (FL05, FL06, IN09, IN10, NC06, NY05, NY06, NY12, MO06) plotted on x-axis. Average maximum temperature (in degrees Celsius) for each environment over the fourth quarter of the growing season plotted on the y-axis. Points are colored by environment. Pearson correlation coefficient (r_p = .37) is shown above the graph. (C) Averaged PR2 values for 9 environments plotted on x-axis. Average minimum temperature (in degrees Celsius) for each environment over the fourth quarter of the growing season plotted on the y-axis. Points are colored by environment. Pearson correlation coefficient (r_p = .47) is shown above the graph.

In order to determine genetic effects on these projections, the calculated values for PR1 through PR7 were used as traits for QTL analysis in each of the 10 environments. Unlike the earlier described PCAs done in environments separately, these PRs are calculated on data from all environments at once. Thus, while the single-environment PCs are distinct between environments, i.e. PC1 in NY05 does not have the same loadings as PC1 in NY06, the PRs are derived from PCs that account for across-environment variance and as such are the same trait for

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mapping in each environment. The mapping analysis yielded at least four QTL for each PR and a total of 44 QTL, eight of which were found in multiple environments and 36 in single environments (Fig 8). Of these multi-environment QTL, 3 are overlapped with the three strongest single element QTL (Mo on Chr1, Cd on Chr 2 and Ni on Chr 9). The Mo and Cd QTL overlap with QTL from different PRs (PRs 3-6 for Mo and PRs 1-5 for Cd) depending on environment while the Ni QTL is only present in PR6 in the NY environments. The presence of PR QTL demonstrate that the covariation in the ionome described by PRs 1-7 results from both environmental and genetic variation. Further investigation is needed to identify the genes underlying PR QTL, their biological roles, and their interaction with specific environmental variables.

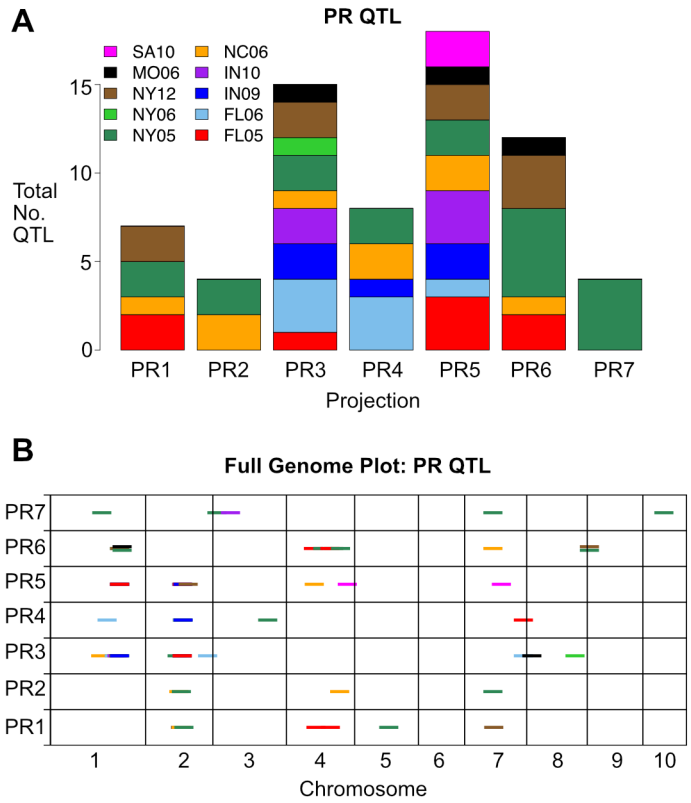


Fig 8. PR QTL in 10 Environments. QTL identified for PCA-derived projection traits (PRs 1-7). (A) Total number of QTL detected for each PR, colored by environment. (B) Significant QTL (loci with LOD scores at or above the 95th percentile of the highest LOD score achieved by

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

running the stepwise QTL mapping algorithm on 1000 random permutation replicates) for PRs 1-7. 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.

Discussion

The results described here demonstrate that the concentrations of elements in the kernels of maize are highly interrelated. Dramatically, and perhaps paradoxically, element concentration is highly heritable within an environment and varied between environments. In agreement with previous studies, we found elemental correlations and QTL that were detected for more than one element. Phosphorous exhibited the greatest number of QTL overlap with other elements, including the cations K and Mg. Phosphorous is a central nutrient in plant development and regulates other elements. In the form of phytate, phosphorous complexes with cations to form phytin in maize seeds [28]. Additional shared QTL included those between Ca and Sr, Mo and Mn, and Zn and Fe. Ca and Sr are chemical analogs while Zn and Fe regulation have been linked at the physiological and molecular level [4,29]. Mo and Mn have roles in protein assimilation and nitrate regulation [30,31] and exhibit a regulatory relationship [18]. Thus, these shared QTL likely reflect the coincident genetic regulation of multiple elements and demonstrate cellular mechanisms to detect and respond to the levels of those elements.

Given the importance of understanding the effects of genetic alteration of multi-element regulatory genes or genetic changes targeted to a single element with pleiotropic effects on other elements, multivariate methods are necessary to create a fuller understanding of the factors underlying ionomic variation. PCA generated variables representative of multiple elements and condensed the covariation between elements without requiring that we know of specify the underlying relationships that led to covariance. QTL mapping with the principal components as traits isolated both loci that had been identified using elements as traits and PC-specific QTL.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The PC-specific QTL likely represent loci with small effects on multiple elements which would not be detected when considering only the concentration of single elements. Candidate genes for these QTL include transporters with multi element specificities or factors within a pathways responding to several elements. Clearly, the inability to detect multiple element loci would leave substantial gaps in our understanding of element regulatory networks, making PCA a useful complement to single element approaches.

The presence of a large number of single-environment QTL is consistent with 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 identified as 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

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

In addition to being a tool for understanding the genetics of multi-element regulation, principal components were also used as a tool to reflect environmental variation in the ionome and genetic by environment interactions. Rather than simply conduct a PCA on all lines, the two-step procedure of generating projections from a PCA on lines replicated within a subset of environments was used to find variables that could describe variation across all 10 environments while minimizing the confounding effect of different sets of lines among all environments. The first two projections 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 PR1 and PR2 to separate growouts by location and, in some cases, different years within a location. Thus, PRs capture the impact of environment on the ionome as a whole. Within the data collected, limited to only maximum and minimum temperature, we observed the strongest correlations for PR1 and PR2 during the fourth quarter of the growing season. Because seed filling occurs in the last quarter of the season, temperature during this time could have a pronounced affect on seed elemental composition. However, given the presence of a low number of data points and a lack of any striking correlations, environmental components of the projections must be largely explained by environmental factors other than temperature. Experiments with more extensive weather and soil data, or carefully manipulated environmental contrasts, are needed to create models with

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

additional covariates and precisely model environmental impacts. The identification of PR QTL indicated that the variation captured by PRs has both environmental and genetic components. These QTL may encompass genes that affect the ionome in an environmentally-responsive manner.

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 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.

Conclusions

Here we have shown 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. Our QTL analysis on single elements found mainly single-

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

environment QTL, indicative of substantial genetic by environment interaction in establishment of the elemental composition of the maize grain. We also demonstrated that treating the ionome as an interrelated set of traits using PCA within environments can identify novel loci. PCA across environments allowed us to derive projections that described both environmental and genetic variation in the ionome. This approach, along with identification of QEI occurring both within a single location over different years and QEI between different locations, demonstrated that gene by environment interactions underlie elemental accumulation in maize kernels.

Methods

Field Growth and Data Collection

Population and field growth. Subsets of the intermated B73 x Mo17 recombinant inbred (IBM) 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 Ukilima, 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. Table S1 includes planting dates and line numbers after outlier removal and genotype matching.

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. Descriptions taken directly are denoted by quotation marks.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). All elements were measured with DRC collision mode off. Before each run, the lens voltage and nebulizer gas flow rate of the ICP-MS were

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 in Davies and Gather (1993). 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

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 whereby the 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 calculate 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

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 sum of squares for genetic or environmental components was 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 sum of squares by the total (genetic plus environmental) sum of squares. 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 *lmfit* followed by *anova* functions to obtain the sum of squares 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

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Principal components analysis. Elements prone to analytical error were removed prior to PC analysis, leaving 16 elements (Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, Se, Rb, Sr, Mo, and Cd) that were used for PCA. 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 and the 10 sets of 16 PCs were used as traits in QTL analysis.

QTL Mapping: principal components. QTL mapping was done in the same manner as described for element phenotypes but with principal components as phenotypes. The mapping procedure was done for each environment separately, with PC line averages for each line in the given environment as phenotypes and line genotypes as input. The stepwise algorithm was used with the same designations as with single-element mapping to produce a QTL model for each PC The PC QTL were compared to element QTL, with unique PC QTL designated as PC QTL at least 25 cM away from any element QTL detected in the environment in which the PC QTL was detected.

QTL by environment analysis: linear model comparison. Linear modeling was used determine instances and strength of QEI using data averaged over years within a three locations (FL, IN, NY). FL05 and FL06 data were combined, averaging for common lines. The same was

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

done with IN09 and IN10 and with NY05, NY06 and NY12. FL, IN, and NY were then used as covariates in QTL analysis. Two QTL models were fit for each phenotype (sample weight, 20 elements, 7 PCs) using the *scanone* function in R/qtl: (1) a model of phenotypic variation with location as an additive and interacting covariate and (2) a model of phenotypic variation with location as only an additive covariate. LOD scores for each marker using model (2) were subtracted from LOD scores for each marker using model (1) to isolate genetic by location variation.

$$V_P = V_G + V_E + V_{GxE} \qquad (1)$$

$$V_P = V_G + V_E \qquad (2)$$

$$V_P = V_{GxE} \qquad (1) - (2)$$

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.

QTL by environment analysis: PCA-derived projections. The 16 most precisely measured elements (Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, Se, Rb, Sr, Mo, and Cd) were used for principal components analysis. The *prcomp* function in R with scale = TRUE was used for PCA

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

on elemental data to perform PCA on only the 16 lines common to 6 of the 10 environments in which we had two replicates per location (FL05, FL06, IN09, IN10, NY05, NY12). PC loadings were extracted from the *prcomp* object and then matrix multiplied with the centered and scaled full dataset matrix (all lines, all environments) to generate whole dataset projections (PRs) of PCs influenced by only lines with multi-environment replicates. The first 7 PRs, derived from the first 7 PCs (98% total variation), were used for correlation analysis with weather variables and stepwise QTL mapping.

Weather data collection and analysis. Weather data for FL05, FL06, IN09, IN10, NC06, NY05, NY06, and NY12 was collected from Climate Data Online (CDO), an archive provided by the National Climatic Data Center (NCDC) through the National Oceanic and Atmospheric Administration website: http://www.ncdc.noaa.gov/cdo-web/. The Climate Data Online Search was used to find Daily Summaries for each day of the growing season from the weather station nearest to the field location. For Florida growouts (FL05, FL06), data was collected from the Homestead General Aviation Airport station. For Indiana growouts (IN09, IN10): West Lafayette 6 NW station. North Carolina (NC06): Clayton station. New York growouts (NY05, NY06, NY12): Aurora Research Farm station. Missouri (MO06): Columbia U of M station. Minimum temperature (in degrees Celsius), and maximum temperature (in degrees Celsius) was 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. GDD was calculated for the entire season and quarterly using the formula GDD = ((Tmax + Tmin)/2) –10. Weather data averages for all environments except for South Africa were tested for correlation with the PR averages from the corresponding 9 environments. The pearson correlation coefficient was calculated for pairs between weather variables and PRs 1-7.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

The authors would like to thank Justin Borevitz and Riyan Cheng, and Tom Juenger for advice on QTL by environment mapping, Karl Broman for invaluable assistance with R/qtl and Sherry Flint-Garcia, Peter Balint-Kurti and Torbert Rocheford for providing seed.

References

1. McDowell SC, Akmakjian G, Sladek C, Mendoza-Cozatl D, Morrissey JB et al. (2013) Elemental Concentrations in the Seed of Mutants and Natural Variants of *Arabidopsis thaliana* Grown under Varying Soil Conditions. PLoS ONE 8: e63014.
2. Cobb JN, DeClerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet 126: 867-887.
3. Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB et al. (2003) Genomic scale profiling of nutrient and trace elements in Arabidopsis thaliana. Nat Biotechnol 21: 1215-1221.
4. Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B et al. (2008) Variation in Molybdenum Content Across Broadly Distributed Populations of *Arabidopsis thaliana* Is Controlled by a Mitochondrial Molybdenum Transporter (*MOT1*). PLoS Genet 4: e1000004.
5. Buescher E, Achberger T, Amusan I, Giannini A, Ochsenfeld C et al. (2010) Natural genetic variation in selected populations of Arabidopsis thaliana is associated with ionomic differences. PLoS One 5: e11081.
6. Baxter I (2009) Ionomics: studying the social network of mineral nutrients. Current

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Opinion in Plant Biology 12: 381-386.

7. Koornneef M, Alonso-Blanco C, Peeters AJM (1997) Genetic approaches in plant physiology. New Phytologist 137: 1-8.

8. Lee M, Sharopova N, Beavis WD, Grant D, Katt M et al. (2002) Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant molecular biology 48: 453-461.

9. Balint-Kurti PJ, Zwonitzer JC, Wisser RJ, Carson ML, Oropeza-Rosas MA et al. (2007) 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 176: 645-657.

10. Dubois PG, Olsefski GT, Flint-Garcia S, Setter TL, Hoekenga OA et al. (2010) Physiological and Genetic Characterization of End-of-Day Far-Red Light Response in Maize Seedlings. Plant Physiol 154: 173-186.

11. Hazen SP, Hawley RM, Davis GL, Henrissat B, Walton JD (2003) Quantitative Trait Loci and Comparative Genomics of Cereal Cell Wall Composition. Plant Physiology 132: 263-271.

12. Lung'aho MG, Mwaniki AM, Szalma SJ, Hart JJ, Rutzke MA et al. (2011) Genetic and physiological analysis of iron biofortification in maize kernels. PLoS One 6: e20429.

13. Ordas B, Malvar RA, Santiago R, Butron A (2010) QTL mapping for Mediterranean corn borer resistance in European flint germplasm using recombinant inbred lines. BMC Genomics 11: 1-10.

14. Zhang N, Gibon Y, Gur A, Chen C, Lepak N et al. (2010) Fine Quantitative Trait Loci Mapping of Carbon and Nitrogen Metabolism Enzyme Activities and Seedling Biomass in

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Maize IBM Mapping Population. Plant Physiology 154: 1753-1765.

15. Shakoor N, Ziegler G, Dilkes BP, Brenton Z, Boyles R et al. (2016) Integration of experiments across diverse environments identifies the genetic determinants of variation in Sorghum bicolor seed element composition. Plant physiology pp-01971.

16. Broman KW, Speed TP (2002) A model selection approach for the identification of quantitative trait loci in experimental crosses. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64: 641-656.

17. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

18. Millikan CR (1948) Antagonism between molybdenum and certain heavy metals in plant nutrition. Nature 161: 528.

19. Bhatia A, Yadav A, Zhu C, Gagneur J, Radhakrishnan A et al. (2014) Yeast Growth Plasticity Is Regulated by Environment-Specific Multi-QTL Interactions. G3 4: 769-777.

20. Leinonen PH, Remington DL, Leppälä J, Savolainen O (2013) Genetic basis of local adaptation and flowering time variation in Arabidopsis lyrata. Molecular ecology 22: 709-723.

21. Nichols KM, Broman KW, Sundin K, Young JM, Wheeler PA et al. (2007) Quantitative Trait Loci x Maternal Cytoplasmic Environment Interaction for Development Rate in Oncorhynchus mykiss. Genetics 175: 335-347.

22. Tétard-Jones C, Kertesz MA, Preziosi RF (2011) 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 366: 1368-1379.

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

23. Ungerer MC, Halldorsdottir SS, Purugganan MD, Mackay TFC (2003) Genotype-Environment Interactions at Quantitative Trait Loci Affecting Inflorescence Development in Arabidopsis thaliana. Genetics 165: 353-365.

24. Des Marais DL, Hernandez KM, Juenger TE (2013) Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annual Review of Ecology, Evolution, and Systematics 44: 5-29.

25. White JG, Zasoski RJ (1999) Mapping soil micronutrients. Field Crops Research 60: 11-26.

26. Godt J, Scheidig F, Grosse-Siestrup C, Esche V, Brandenburg P et al. (2006) The toxicity of cadmium and resulting hazards for human health. J Occup Med Toxicol 1: 22.

27. United States Department of Agriculture GAIN (2014) China's Maximum Levels of Contaminants in Food. GAIN report CH14058. Available: http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Maximum%20Levels%20of%20Contaminants%20in%20Foods%20_Beijing_China%20-%20Peoples%20Republic%20of_12-11-2014.pdf via the Internet. Accessed x.

28. López-Arredondo DL, Leyva-González MA, González-Morales SI, López-Bucio J, Herrera-Estrella L (2014) Phosphate nutrition: improving low-phosphate tolerance in crops. Annual Review of Plant Biology 65: 95-123.

29. Lin Y-F, Liang H-M, Yang S-Y, Boch A, Clemens S et al. (2009) Arabidopsis IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. New Phytologist 182: 392-404.

30. Mulder EG (1948) Importance of molybdenum in the nitrogen metabolism of microorganisms and higher plants. Plant and Soil 1: 94-119.

31. Mulder EG, Gerretsen FC (1952) Soil manganese in relation to plant growth. Adv Agron 4:

bioRxiv preprint doi: https://doi.org/10.1101/048173; this version posted April 13, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

221-277.

Supporting Information

S1 Fig. 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.

S2 Fig. Variances of Principal Components from PCA on Common Lines in Six Environments. Eigenvalues (amount of variation explained) for each PC are shown on the y-axis. PC1-7 account for 98% of total variance.

S1 Table. Growout Information.

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

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