bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license.

Genome-wide association study reveals an independent genetic basis of zinc and cadmium concentrations in fresh sweet corn kernels

Matheus Baseggio[*,1], Matthew Murray[†], Di Wu[*], Gregory Ziegler[‡], Nicholas Kaczmar[*,2], James Chamness[*,3], John P. Hamilton[§], C. Robin Buell[§], Olena K. Vatamaniuk[**], Edward S. Buckler[*,††,‡‡], Margaret E. Smith[*], Ivan Baxter[‡], William F. Tracy[†], and Michael A. Gore[*,4]

[*]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA

[†]Department of Agronomy, University of Wisconsin-Madison, Madison, WI, 53706 USA

[‡]Donald Danforth Plant Science Center, St. Louis, MO 63132, USA

[§]Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

[**]Soil and Crop Sciences Section, Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853 USA

[††]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853 USA

[‡‡]US Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, NY 14853 USA.

[1]Present address: Seneca Foods Corporation, 1201 N. 4th Street, LeSueur, MN 56058, USA; [2]Present address: Horticulture Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853 USA; [3]Present address: Department of Genetics, Cell Biology, and Development, University of Minnesota, Saint Paul, MN 55108, USA; [4]Corresponding author: 358 Plant Science Building, Cornell University, Ithaca, NY 14853. Email: mag87@cornell.edu

Running head: Sweet corn elemental kernel phenotypes

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license.

Abstract

Despite being one of the most consumed vegetables in the United States, the elemental profile of sweet corn (*Zea mays* L.) is limited in its dietary contributions. To address this through genetic improvement, a genome-wide association study was conducted for the concentrations of 15 elements in fresh kernels of a sweet corn association panel. In concordance with mapping results from mature maize kernels, we detected a probable pleiotropic association of zinc and iron concentrations with *nicotianamine synthase5* (*nas5*), which purportedly encodes an enzyme involved in synthesis of the metal chelator nicotianamine. Additionally, a pervasive association signal was identified for cadmium concentration within a recombination suppressed region on chromosome 2. The likely causal gene underlying this signal was *heavy metal ATPase 3* (*hma3*), whose counterpart in rice, *OsHMA3*, mediates vacuolar sequestration of cadmium and zinc in roots, whereby regulating zinc homeostasis and cadmium accumulation in grains. Consistent with transgenic studies in rice, we detected an association of *hma3* with cadmium but not zinc accumulation in fresh kernels. This finding implies that selection for low cadmium will not affect zinc levels in fresh kernels. Although less resolved association signals were detected for boron, nickel, and calcium, all 15 elements were shown to have moderate predictive abilities via whole-genome prediction. Collectively, these results help improve our genomics-assisted breeding efforts centered on improving the elemental profile of fresh sweet corn kernels.

Keywords: genome-wide association study, whole-genome prediction, elements, kernels, sweet corn

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Introduction

As with all multicellular organisms, the concentration and distribution of elements in tissues and organs influence growth and development over the plant life cycle. At least 16 elements (boron, calcium, carbon, chlorine, copper, hydrogen, iron, magnesium, manganese, molybdenum, nitrogen, oxygen, phosphorus, potassium, sulfur, and zinc) are considered essential for higher plant species, with an additional four elements (cobalt, silicon, nickel, and sodium) essential for a subset of higher plants (Mengel and Kirkby 2012). In plants, the need for and concentration of macroelements (carbon, hydrogen, oxygen, nitrogen, phosphorus, sulfur, potassium, calcium, magnesium) (Hawkesford *et al.* 2012) are relatively greater than for microelements (iron, manganese, copper, zinc, nickel, molybdenum, boron, chlorine, and cobalt) (Broadley *et al.* 2012). Additionally, nonessential heavy metals such as cadmium, chromium, and lead that lack involvement in normal physiological functions can accumulate to toxic concentrations in plants, penetrating the food chain and posing a threat to human health (Singh *et al.* 2016).

Not unlike plants, humans can suffer a range of adverse health effects from the excess or deficiency of essential and nonessential elements. Of all the micronutrients, deficiency is most prevalent for iron, with more than two billion people affected worldwide (Viteri 1998). Of similar scale, nearly two billion people are estimated to suffer from dietary zinc deficiency throughout developing nations (Prasad 2014). Given that dietary substances such as phytate in cereal grains bind zinc and iron and inhibit their absorption, zinc and iron deficiencies resulting from low bioavailability could coincide (Sandstead and Smith 1996; Lönnerdal 2000). Although severe dietary micronutrient deficiency is far less prevalent in developed nations, approximately 10 million people are iron deficient in the U.S. (Miller 2013). In the U.S., the average daily

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

intake of iron by most premenopausal (12 mg d^{-1}) and pregnant (15 mg d^{-1}) women is 6 and 12 mg d^{-1} less than its recommended daily allowance (RDA), respectively (Institute of Medicine 2001; Linus Pauling Institute 2016). Additionally, the zinc RDA for adult women and men is 11 mg d^{-1} and 8 mg d^{-1}, with elderly in the U.S. at higher risk for mild zinc deficiency (Mocchegiani *et al.* 2013; Linus Pauling Institute 2019).

Crop biofortification via genomics-assisted breeding and genetic engineering has emerged as an attractive approach for nutritional enhancement of crops and the generation of new varieties with a high density of iron and zinc in edible plant tissues (Murgia *et al.* 2012; Bhullar and Gruissem 2013; Hirschi 2020). It is increasingly recognized that plant membrane transporters and metal chelators are among the key targets for increasing mineral nutrient density in plant tissues (Waters and Sankaran 2011; Schroeder *et al.* 2013). However, in many cases, transporters that facilitate the accumulation of iron and zinc are multispecific and can mediate the uptake and internal transport of nonessential and potentially highly toxic heavy metals including cadmium (Waters and Sankaran 2011; Schroeder *et al.* 2013; Khan *et al.* 2014). Therefore, efforts to increase the concentration of iron and zinc in grains of cereals could also increase the concentration of cadmium. This poses a serious threat to food security, especially if crops are grown on soils either contaminated with cadmium or low in microelements, particularly iron (Waters and Sankaran 2011; Schroeder *et al.* 2013; Khan *et al.* 2014).

Vegetative and seed tissues of fruits and vegetables are important dietary sources of essential and nonessential elements for humans to meet their daily nutrient needs. Given that sweet corn is the third most consumed vegetable in the U.S. (USDA-NASS 2018), the elemental profile of fresh sweet corn kernels is an important consideration for human health and nutrition. Although not a major contribution to the RDA of essential microelements, the consumption of

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

100 g of uncooked, yellow sweet corn (medium-sized ear) provides 0.52 mg of iron and 0.46 mg of zinc (USDA-ARS 2019), but the bioavailable amount is expected to be no more than a quarter of the total of each element (Bouis and Welch 2010). Therefore, there exists a tremendous opportunity to improve the elemental profile of fresh sweet corn kernels through genomics-assisted breeding, but this first requires an understanding of the phenotypic variability and genetic control of zinc, iron, and other elements. Considerable heritable variation exists for elemental concentrations in physiologically mature grain of diverse maize panels (Ziegler *et al.* 2017; Wu *et al.* 2021 in review), but a comparable level of genetic understanding is severely lacking for immature kernels (fresh-eating stage) of diverse sweet corn germplasm.

Complex physiological and genetic networks coordinate elemental uptake, transport, and accumulation in plants, and these processes are responsive to the environment in which plants are grown (Reviewed in Baxter 2009). As noted above, metal transporters and chelators are central to the accumulation of elements in seed whether arising from direct uptake by roots or remobilization from vegetative tissues (Waters and Sankaran 2011). Importantly, plant transporters and chelators can have broad metal specificity (Axelsen and Palmgren 2001), which partially explains why certain elements are genetically interconnected (Reviewed in Baxter 2015). Taken together, the influence of the environment and possible multi-element specificity of metal transporters and chelators must be considered when elucidating the physiological and genetic mechanisms that underpin accumulation of elements in fresh sweet corn kernels.

In the genomes of Arabidopsis, maize, rice, and other plant species, gene families have been identified for metal transporters and chelators including but not limited to HEAVY-METAL ATPASE (HMA), OLIGOPEPTIDE TRANSPORTERS (OPTs) and their subfamily of YELLOW STRIPE-LIKE (YSL), ZINC-REGULATED TRANSPORTER (ZRT)/IRON-

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

REGULATED TRANSPORTER (IRT)-LIKE PROTEIN (ZIP) and NICOTIANAMINE SYNTHASE (NAS) (Whitt *et al.* 2020). As it relates to maize, *yellow stripe1* (*ys1*) and *ys3* encode proteins that have been functionally shown to transport iron when associated with a strong metal-ligand, nicotianamine or its derivative phytosiderophores such as mugineic acid and deoxymugenic acid (Von Wiren *et al.* 1994; Chan-Rodriguez and Walker 2018), whereas the proteins encoded by *ysl2* and *zip5* have been functionally implicated in the accumulation of zinc and iron in grain (Li *et al.* 2019; Zang *et al.* 2020).

Of the metal chelators, nicotianamine, is considered among the main contributors to zinc and iron accumulation in grain (Curie et al., 2009). Nicotianamine is synthesized from *S*-adenosyl-methionine by NAS enzymes that are encoded by a multigene family in Arabidopsis, rice, maize, and barley (Higuchi *et al.* 1999; Klatte *et al.* 2009; Zhou *et al.* 2013). The nine *nas* genes that were identified in maize are grouped into two classes based on a phylogenetic analysis; these groups also differ in tissue specificity of their expression, responsiveness to element availability, and are suggested to function in either element uptake into maize roots or internal transport and homeostasis (Zhou *et al.* 2013). Despite these advancements, the vast majority of members from this gene family and others in maize have not been deeply characterized at the functional level, thus a wide gap in the knowledge base remains for the key transporters and chelators controlling the content and composition of elements in tissues and organs.

Explaining and predicting the quantitative variation of phenotypes is a major challenge in crop plants, but there has been notable recent progress for maize grain elemental phenotypes (Ziegler *et al.* 2017; Hindu *et al.* 2018; Wu *et al.* 2021 in review). In the U.S. maize nested association mapping (NAM) panel, joint-linkage analysis and genome-wide association studies

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

(GWAS) were used to identify six strong candidate genes for the concentrations of manganese, molybdenum, phosphorus, or rubidium in physiologically mature grain (Ziegler *et al.* 2017). Through the implementation of GWAS in the maize Ames panel, Wu *et al.* (2021 in review) resolved several loci previously identified to control variation for copper, iron, manganese, molybdenum, and/or zinc in mature grain from the U.S. NAM panel, which resulted in the identification of two metal chelator and five metal transporter candidate genes. Additionally, the authors detected novel candidate gene loci for boron and nickel grain concentrations. Whole-genome prediction (WGP) models have been found to be moderately predictive of elemental concentrations in mature grain of tropical maize populations (zinc) (Guo *et al.* 2020; Mageto *et al.* 2020) and the Ames panel (boron, calcium, copper, iron, potassium, magnesium, manganese, molybdenum, nickel, phosphorus, and zinc) (Wu *et al.* 2021 in review). Despite this progress with mature grain, the genotype-phenotype map of elemental concentrations in fresh sweet corn kernels is completely nonexistent, thus there exists tremendous opportunities for studying the quantitative genetics of these nutritionally relevant phenotypes.

In this study, we used a sweet corn association panel for the genetic dissection and prediction of quantitative variation of 15 elements in fresh sweet corn kernels. The three major objectives of our study were to (i) evaluate variability and heritability of elemental fresh kernel phenotypes within and across field locations, (ii) employ GWAS to identify candidate genes associated with the levels of elements in fresh kernels, and (iii) assess the predictive abilities of WGP models as an evaluation of the potential that genomic selection has for the genetic improvement of elemental fresh kernel phenotypes important to human nutrition and health.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Materials and Methods

Plant Materials and Experimental Design

In two consecutive field seasons (2014-15), a sweet corn association panel of inbred lines representing the genetic diversity of temperate U.S. breeding programs (Baseggio *et al.* 2019) was evaluated at Cornell University's Musgrave Research Farm in Aurora, NY, on a Lima silt loam (fine-loamy, mixed, semiactive, mesic Oxyaquic Hapludalfs) and University of Wisconsin's West Madison Research Station in Verona, WI, on a Plano silt loam (fine-silty, mixed, superactive, mesic Typic Argiudolls). The panel consists of *sugary1* (*su1*), *sugary1:sugary enhancer1* (*su1se1*), *shrunken2* (*sh2*), *sugary1:shrunken2* (*su1sh2*), *brittle2* (*bt2*), and *amylose-extender:dull:waxy* (*aeduwx*) lines that are homozygous for endosperm mutations that cause deficiences in starch biosynthesis. Additionally, there were 20 non-sweet corn inbred lines and four repeated check sweet corn inbred lines included in the experiment. In each of the four environments (location × year combination), the experiment was arranged as an augmented incomplete block design grown as a single replicate as previously described by Baseggio *et al.* (2019). Briefly, the lines were separated into three sets according to their plant height, with each set having incomplete blocks. Each incomplete block of 20 experimental lines was augmented with the random placement of two height-specific check lines (We05407 and W5579, W5579 and Ia5125, or Ia5125 and IL125b). In both field locations, experimental units were one-row plots of different lengths. Plots were 3.05 m long in NY and 3.50 m long in WI. Both locations had an inter-row spacing of 0.76 m, with a 0.91 m alley at the end of each plot. In NY, 25 kernels were planted per plot and thinned to 12 plants per plot. In WI, 12 kernels were planted in each plot, but plots were not thinned.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

In all environments, multiple plants per plot were selfed pollinated, with two selfed ears collected by hand from each harvestable plot at 400 growing degree days (i.e., immature kernel stage at approximately 21 days after pollination) as earlier described (Baseggio *et al.* 2019). Immediately upon fresh harvest, the entirety of each dehusked ear was directly frozen in liquid nitrogen, followed by hand shelling of frozen kernels. To generate a representative composite kernel sample for each harvested plot, frozen kernels were equally sampled at random from both ears, bulked, and stored in 15 mL Falcon tubes at -80°C until lyophilization. A combined set of 1,524 plot samples from across all environments, with each sample consisting of three lyophilized kernels, was shipped to the Donald Danforth Plant Science Center (St. Louis, MO) for elemental analysis.

Phenotypic Data Analysis

The determination of elemental concentration by an inductively coupled plasma mass spectrometer (ICP-MS) on a single-kernel basis was conducted as previously described in Baxter *et al.* (2014). In short, each individual unground kernel was robotically weighed, digested in concentrated nitric acid, and measured for concentrations of aluminum, arsenic, boron, cadmium, calcium, cobalt, copper, iron, magnesium, manganese, molybdenum, nickel, phosphorus, potassium, rubidium, selenium, sodium, strontium, sulfur, and zinc with a PerkinElmer NexION 350D ICP-MS. Of these 20 elements, aluminum, arsenic, cobalt, selenium, and sodium were not further considered because their measured concentrations were at trace levels, vulnerable to contamination in the course of sample processing, and/or sensitive to interference from other sample matrix constituents (Ziegler *et al.* 2013). To limit the influence of extreme analytical outliers that could negatively affect the accurate estimation of variance components when initially fitting a mixed linear model to the raw data, the method of Davies and Gather (1993)

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

was implemented similarly to its use in Baxter *et al.* (2014) to remove raw concentration values with greater than a conservative threshold of 15 median absolute deviations from the median concentration for a given element within each environment. If less than 1% of the values for a given element were negative, these negative values were set to missing.

The preliminarily processed raw ICP-MS dataset was more robustly screened for significant outliers by fitting a mixed linear model that allowed for genetic effects to be separately estimated from field design effects, following the procedure described in Wolfinger *et al.* (1997). The fitted mixed linear model was similar to that used by Baseggio *et al.* (2019) for the same experimental field design, with the notable exception that the model used in this study included a term to estimate within-plot kernel sample variance. This allowed for the removal of individual outlier measurements. For each elemental phenotype**,** the full model was fitted in ASReml-R version 3.0 (Gilmour *et al.* 2009) across locations (all four environments) or for each location separately (two environments, NY; or two environments, WI) as follows:

$$Y_{ijklmnopq} = \mu + check_i + env_j + set(env)_{jk} + block(set \times env)_{jkl} + genotype_m + (genotype \times env)_{jm} + ICP.run_n + sample_o + row(env)_{jp} + col(env)_{jq} + \varepsilon_{ijklmnopq} \quad [1]$$

in which $Y_{ijklmnopq}$ is an individual phenotypic observation, μ is the grand mean, $check_i$ is the fixed effect for the *i*th check, env_j is the effect of the *j*th environment, $set(env)_{jk}$ is the effect of the *k*th set within the *j*th environment, block$(set \times env)_{jkl}$ is the effect of the *l*th incomplete block within the *k*th set within the *j*th environment, $genotype_m$ is the effect of the *m*th experimental genotype (non-check line), $(genotype \times env)_{jm}$ is the effect of the interaction between the *m*th genotype and *j*th environment, $ICP.run_n$ is the laboratory effect of the *n*th ICP run, $sample_o$ is the *o*th kernel sample, $row(env)_{jp}$ is the effect of the *p*th plot grid row within the *j*th environment, $col(env)_{jq}$ is the effect of the *q*th plot grid column within the *j*th environment, and $\varepsilon_{ijklmnopq}$ is the

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

heterogeneous residual error effect within each environment with a first order autoregressive correlation structure among plot residuals in the row and column directions. With the exception of the grand mean and check term, all terms were modeled as random effects. The Kenward-Roger approximation (Kenward and Roger 1997) was used to calculate degrees of freedom. Studentized deleted residuals (Neter *et al.* 1996) obtained from these mixed linear models were used to detect significant outliers for each phenotype after a Bonferroni correction ($\alpha = 0.05$).

To generate BLUP values for each elemental phenotype, an iterative mixed linear model fitting procedure was conducted on the outlier-screened phenotypic dataset in ASReml-R version 3.0 (Gilmour *et al.* 2009) with the full model across locations or for each location separately. Model terms fitted as random effects including the autoregressive correlations were tested with likelihood ratio tests (Littell *et al.* 2006), followed by the removal of terms from the model that were not significant at $\alpha = 0.05$. The significance of main random effects are reported in Table S1. Additionally, the first order autoregressive correlation structure was statistically significant for all phenotypes. For each elemental phenotype, the final, best fitted model was used to generate a best linear unbiased predictor (BLUP) for each inbred line. The generated BLUP values were filtered to remove non-sweet corn lines, as well as sweet corn lines with the infrequent *aeduwx* or *bt2* endosperm mutations and those without available SNP marker data. This resulted in 401 sweet corn lines with more prevalent endosperm mutations [*su1*, *su1se1* (classified as *su1* for this study due to lack of informative marker genotypes), *sh2*, and *su1sh2*] that had BLUP values for elemental phenotypes across and within locations.

With variance component estimates from each best fitted model, heritability on a line-mean basis was calculated for each elemental phenotype across locations and separately for each location as previously described (Lynch and Walsh 1998; Holland *et al.* 2003; Hung *et al.* 2012).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Pearson's correlation coefficient (*r*) was used to assess the degree of association between the BLUP values of paired phenotypes. Pairwise correlations were calculated and their significance tested at $\alpha = 0.05$ with the method 'pearson' from the function 'cor.test' in R version 3.6.1 (R Core Team 2019).

SNP Marker Genotyping

The sweet corn inbred association panel was sequenced via the genotyping-by-sequencing (GBS) procedure of Elshire *et al.* (2011) with *Ape*KI at the Cornell Biotechnology Resource Center (Cornell University, Ithaca, NY, USA) as previously described (Baseggio *et al.* 2019). The procedure of Baseggio *et al.* (2020) for SNP calling, filtering, and imputing missing genotypes was used to construct a SNP marker dataset for the genetic dissection and prediction of fresh kernel elemental phenotypes. In brief, the raw GBS sequencing data from Baseggio *et al.* (2019) was processed through the production pipeline in TASSEL 5 GBSv1 with the ZeaGBSv2.7 Production TagsOnPhysicalMap file to call SNPs at 955,690 loci in B73 RefGen_v2 coordinates (Glaubitz *et al.* 2014). These raw SNP genotype calls were merged with those of 19 sweet corn inbred lines from Romay et al. (2013) that were not included in the Baseggio *et al.* (2019) GBS dataset, allowing for the assemblage of raw SNP calls for all 401 sweet corn inbred lines with BLUP values.

The combined raw dataset was initially filtered by keeping only biallelic SNPs with a call rate > 10% and eliminating SNPs with a minor allele scored in only a single individual (singletons and doubletons). Given the potential to have resulted from paralogous alignments, we set heterozygous genotype calls with an allele balance score (lowest allele read depth/total read depth) < 0.3 to missing. If multiple GBS samples existed for an inbred line, SNP genotype calls from samples with the same accession number/identifier were merged and discordant genotype

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

calls set to missing if identical-by-state (IBS) values for all within-line sample comparisons were > 0.99, following the conservative IBS threshold set by Romay *et al.* (2013). A single GBS sample with the highest SNP call rate was chosen to represent an inbred line if all pairwise IBS values were less than 0.99.

The FILLIN haplotype-based imputation strategy of Swarts *et al.* (2014) was used to impute missing SNP genotypes to near completeness based on a set of maize haplotype donors with a window size of 4 kb. To improve the quality of the imputed dataset, we filtered SNPs in TASSEL 5 version 20190321 to remove those with a call rate < 70% (residual missing genotype data are expected for the haplotype-based imputation method of FILLIN; Swarts *et al.*, 2014), a minor allele frequency < 5%, heterozygosity > 10%, coefficient of panmixia< 80%, or a mean read depth > 15. To uplift the genome coordinates of retained SNPs to B73 v4, the Vmatch software (Kurtz 2003) was used to align the 101 bp context sequence of each SNP to the B73 RefGen_v4 reference genome, resulting in 147,762 high-quality SNP markers scored on the 401 sweet corn inbred lines.

Genome-Wide Association Study

A GWAS was conducted across and within locations to identify SNP markers significantly associated with each elemental phenotype following the methods of Baseggio *et al.* (2020) with minor modifications. In short, the Box-Cox power transformation (Box and Cox 1964) was used with an intercept-only model to choose the optimal value of convenient lambda (-2 to +2, 0.5 increments) (Table S2) for transforming the BLUP values of each elemental phenotype to lessen heteroscedasticity and non-normality of the residuals with the MASS package in R version 3.6.1 (R Core Team 2019). For each elemental phenotype, a mixed linear model (Yu *et al.* 2006; Zhang *et al.* 2010) that accounted for population structure and unequal relatedness with principal

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

components (PCs) and a genomic relationship matrix (GRM; kinship) was used to test for an association between each of the 147,762 SNPs and transformed BLUP values in GEMMA software version 0.97 (Zhou and Stephens 2014). In the R package GAPIT version 2017.08.18 (Lipka *et al.* 2012), 10,773 unimputed genome-wide SNPs (call rate >90%, MAF >5%, heterozygosity <10%, coefficient of panmixia >80%, and mean read depth <15) subsampled from the complete marker dataset were used to calculate PCs with the prcomp function and the kinship matrix with VanRaden's method 1 (VanRaden 2008). The conservative imputation of residual missing SNP genotypes as heterozygous in both marker datasets was conducted in GAPIT.

The Bayesian information criterion (BIC) (Schwarz 1978) was used to ascertain the optimal number of PCs to incorporate in the mixed linear model. Given that the uneven distribution of elements among seed tissues such as the endosperm and embryo (Lombi *et al.* 2009, 2011; Pongrac *et al.* 2013; Baxter *et al.* 2014; Cheah *et al.* 2019) could potentially lead to spurious associations with *su1* and *sh2* as shown for tocotrienols and certain carotenoids (Baseggio *et al.* 2019, 2020), endosperm mutation type (*su1*, *sh2*, or *su1sh2*) was also tested with the BIC for inclusion as a covariate in the model. Of the 401 inbred lines, 384 lines had endosperm mutation type scores available from Baseggio *et al.* (2019), whereas endosperm mutation type for each of the remaining 17 lines without visual scores was predicted with the identical optimal marker-based classification models and 1000 kb marker datasets for the *su1* and *sh2* loci from Baseggio *et al.* (2019).

To approximate the amount of phenotypic variation explained by a significantly associated SNP, we calculated the difference between the likelihood-ratio-based R^2 statistic (R^2_{LR}) of Sun *et al.* (2010) from a mixed linear model with or without the given SNP, following

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Baseggio *et al.* (2020). The false-discovery rate (FDR) was controlled at 5% by adjusting the *P*-values (Wald test) of SNPs tested in GEMMA using the Benjamini–Hochberg multiple test correction (Benjamini and Hochberg 1995) with the 'p.adjust' function in R version 3.6.1 (R Core Team 2019). Given the distance to which genome-wide linkage disequilibrium (LD) decays to background levels in this association panel (Baseggio *et al.* 2019) and to account for the possibility of distant regulatory elements (enhancers), candidate gene searches were limited to ± 250 kb of the physical position of SNP markers significantly associated with an elemental phenotype. For each most plausible candidate gene, we used BLASTP to identify the top three unique best hits (E-values < 1) in Arabidopsis (Columbia-0 ecotype) and rice (*Oryza sativa* L. ssp. Japonica cv. 'Nipponbare') using default parameters at the TAIR (https://www.arabidopsis.org) and RAP-DB (https://rapdb.dna.affrc.go.jp) databases, respectively. The across-location (All Locs: New York, Florida, North Carolina, and Puerto Rico) results from JL analysis and GWAS of grain elemental phenotypes in the U.S. NAM panel (Ziegler *et al.* 2017) were integrated with the physical (bp) positions of GWAS signals from our study in B73 RefGen_AGPv4 coordinates following the approach of Wu *et al.* (2021 in review).

The multi-locus mixed-model (MLMM) approach of Segura *et al.* (2012) that sequentially adds significant markers as covariates in the model was used to better clarify significant association signals with underlying large-effect loci at the level of an individual chromosome as previously described (Lipka *et al.* 2013). The optimal model was selected with the extended BIC (Chen and Chen 2008). To further assess the extent of statistical control for large-effect loci, GWAS was reconducted by including MLMM-selected SNPs as fixed effects (covariates) in mixed linear models fitted in GEMMA.

Linkage disequilibrium

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

The local patterns of LD surrounding significantly associated loci were investigated by estimating pairwise LD between SNPs with the squared allele-frequency correlation (r^2) method of Hill and Weir (1988) in TASSEL 5 version 20190627 (Bradbury *et al.* 2007). The marker dataset used for estimation of LD consisted of the 147,762 SNPs without imputation of the post-FILLIN residual missing SNP genotypes to heterozygotes.

Whole-genome prediction

A univariate genomic best linear unbiased prediction (GBLUP) model (Bradbury *et al.* 2007; VanRaden 2008) was used to evaluate whole-genome prediction (WGP) on the transformed across-location BLUP values of the 15 elemental phenotypes as previously described by Baseggio *et al.* (2020). In short, the 401 line x 147,762 SNP genotype matrix with post-FILLIN missing data imputed as a heterozygous genotype was used to construct a GRM with method 1 from VanRaden (2008) in GAPIT version 2017.08.18 (Lipka *et al.* 2012). Next, the constructed GRM was modeled as a random effect to predict each individual elemental phenotype with the function 'emmreml' in version 3.1 of the R package EMMREML (Akdemir and Okeke 2015). Through the implementation of a five-fold cross-validation scheme conducted 50 times for each elemental phenotype, the predictive ability of a phenotype was calculated as the mean Pearson's correlation between transformed BLUP (observed) and genomic estimated breeding values (predicted). Each fold was representative of genotype frequencies for endosperm mutants (*su1*, *sh2*, and *su1sh2*) observed in the whole association population. Endosperm mutation type (*su1*, *sh2*, or *su1sh2*) was also evaluated as a covariate in prediction models, with the same cross-validation folds used across models with or without the covariate for endosperm mutation type.

Data availability

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

All raw GBS sequencing data are available from the National Center of Biotechnology Information Sequence Read Archive under accession number SRP154923 and in BioProject under accession PRJNA482446. The ZeaGBSv2.7 Production TagsOnPhysicalMap file (AllZeaGBSv2.7_ProdTOPM_20130605.topm.h5) for calling SNPs, the raw SNP genotype data in B73 AGPv2 coordinates (ZeaGBSv27_publicSamples_rawGenos_AGPv2-150114.h5) for the 19 sweet corn lines of Romay *et al.* (2013), and the maize haplotype donor file (AllZeaGBSv2.7impV5_AnonDonors4k.tar.gz) for imputing missing genotypes are on CyVerse (https://datacommons.cyverse.org/browse/iplant/home/shared/panzea/genotypes/GBS/v27). The BLUP values of the 15 elemental phenotypes and the FILLIN imputed SNP genotype calls in B73 AGPv4 coordinates for the 401 inbred lines are available at CyVerse: (https://datacommons.cyverse.org/browse/iplant/home/shared/GoreLab/dataFromPubs/Baseggio_SweetcornElement_2021). The Supplemental Figures and Tables are available at CyVerse: (https://datacommons.cyverse.org/browse/iplant/home/shared/GoreLab/dataFromPubs/Baseggio_SweetcornElement_2021/BioRxivSupplementalInformation).

Results

Phenotypic variation

The extent of phenotypic variation for 15 elements in fresh kernels as quantified by ICP-MS was evaluated in an association panel of 401 sweet corn inbred lines that was grown in two field locations (Verona, WI; and Aurora, NY) in 2014 and 2015. Of the five macroelements studied, potassium, phosphorus, sulfur, and magnesium had average concentrations greater than 1,000 µg g^{-1}, whereas calcium had an average concentration of nearly 60 µg g^{-1} (Table 1). Strikingly, average concentrations ranged from 0.012 (cadmium) to 24.52 (zinc) µg g^{-1} for the 10 microelements. Even though cadmium had the lowest mean concentration, it had a 12.11-fold

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

range in variation (maximum BLUP value divided by the minimum BLUP value), whereas the other 14 elements covered a 1.34- to 3.66-fold range in variation. When separating inbred lines according to their endosperm mutation type (Table 2), copper, iron, manganese, and potassium were found to be at significantly ($P < 0.001$) greater concentrations in the *sh2* ($n = 78$) group relative to the *su1* ($n = 301$) group.

Table 1. Means and ranges for untransformed best linear unbiased predictors (BLUPs) of 15 fresh kernel elemental phenotypes evaluated in the sweet corn association panel and estimated heritability ($\hat{h}_l^2$) on a line-mean basis across two years and two locations.

Trait	Lines	BLUPs			$\hat{h}_l^2$
		Mean	SD[a]	Range	
		——— μg g^{-1} dry weight ———			
Boron	399	1.89	0.12	1.62-2.35	0.15
Cadmium	401	0.012	0.006	0.003-0.041	0.83
Calcium	392	57.91	10.76	34.36-106.08	0.48
Copper	401	3.25	0.76	1.61-5.90	0.81
Iron	401	18.55	2.08	13.66-28.11	0.64
Magnesium	401	1268.23	71.83	1088.97-1574.82	0.64
Manganese	401	8.52	1.21	5.33-13.13	0.86
Molybdenum	400	0.301	0.040	0.209-0.438	0.88
Nickel	401	0.161	0.032	0.096-0.312	0.64
Phosphorus	401	3186.06	165.14	2713.77-3639.37	0.76
Potassium	401	8089.79	473.62	6558.37-9765.09	0.76
Rubidium	401	2.72	0.25	2.22-3.72	0.58
Strontium	401	0.273	0.044	0.177-0.503	0.60
Sulfur	401	1451.63	73.60	1258.22-1712.46	0.40
Zinc	401	24.52	2.26	18.58-32.10	0.75

[a]Standard deviation of the BLUPs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Table 2. Estimated effects of endosperm mutation type from untransformed best linear unbiased predictors of 15 fresh kernel elemental phenotypes across two years and two locations.

Trait	***su1***[a]	***sh2***	***su1sh2***	***P*-value**[b]
	——— μg g^{-1} dry weight ———			
Boron	1.89	1.88	1.93	0.147
Cadmium	0.011 a	0.013 a	0.013 a	**0.013**
Calcium	57.79	57.05	62.48	0.105
Copper	3.12 b	3.57 a	3.93 a	**<0.001**
Iron	18.31 b	19.16 a	19.68 a	**<0.001**
Magnesium	1262.99 b	1275.52 ab	1314.02 a	**0.003**
Manganese	8.39 b	8.77 a	9.44 a	**<0.001**
Molybdenum	0.301	0.302	0.294	0.732
Nickel	0.158 a	0.166 a	0.173 a	**0.035**
Phosphorus	3173.18 b	3206.30 ab	3290.45 a	**0.003**
Potassium	8027.32 c	8216.81 b	8494.14 a	**<0.001**
Rubidium	2.72	2.72	2.72	0.999
Strontium	0.271 b	0.274 b	0.300 a	**0.010**
Sulfur	1444.09 b	1462.47 b	1516.39 a	**<0.001**
Zinc	24.43	24.64	25.24	0.237

[a]Sweet corn lines grouped by endosperm mutation type having labels with the same letter are not significantly different according to the Tukey-Kramer honest significant difference test ($P < 0.05$). The test was only performed for phenotypes that had a significant F-test.

[b]P-value from one-way analysis of variance (ANOVA) F-test for the endosperm mutation type effect. Bolded P-value indicates a statistically significant difference between two or more endosperm mutation type groups ($P < 0.05$).

Implying common genetic control (Baxter 2015), shared chemical and physiological properties (Marschner 2012), or storage with phytic acid (Maathuis 2009), element pairs with strong positive correlations ($r > 0.50$; $P < 0.01$) across locations (Figure S1) were as follows: strontium/calcium, magnesium/phosphorus, zinc/iron, and phosphorus/zinc. Suggestive of a distinct genetic architecture, molybdenum was the element most weakly correlated with other elements across locations (Figure S1), having a significant but very weak positive correlation ($r = 0.12$; $P = 0.02$) with zinc alone. With the exception of boron ($\hat{h}_l^2 = 0.15$), the across-location heritability estimates (Table 1) for the elemental phenotypes ranged from 0.40 (sulfur) to 0.88 (molybdenum). The genotype term was significant for all 15 elements, while the G×E interaction term was significant for boron, copper, iron, magnesium, and rubidium (Figure S2 and Table

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

S1).

When investigating phenotypic variation between locations, we found that only the average concentration of sulfur was not significantly different between locations (Figure S3). Indicative of phenotypes with a range of responsiveness to the environment, the correlation (r) of elemental trait BLUPs between locations ranged from 0.08 (boron) to 0.68 (copper), with an average correlation of 0.42 (Figure S4). Despite the mostly moderate correlations between locations, within-location heritability estimates (Table S3) were comparable to those estimated across locations (Table 1) and strongly correlated (r = 0.81) between the NY and WI locations. Altogether, our findings suggest that there is value in exploring the genetic dissection of elemental fresh kernel phenotypes across- and within-locations.

Genome-Wide Association Study

We investigated the genetic basis of natural variation for the concentration of 15 elements in fresh kernels from the sweet corn association panel of 401 inbred lines that had been evaluated in four environments (two years × two locations) and scored with 147,762 genome-wide SNP markers. Through an across-location GWAS conducted with a unified mixed linear model that accounted for population structure, relatedness, and endosperm mutation type, 220 unique SNPs were found to associate with one of three elements (cadmium, zinc, or boron) at a genome-wide FDR of 5% (Table S4). Significant association signals were only found on chromosomes 1 (boron), 2 (cadmium), and 7 (zinc), with the exception of a single SNP associated with cadmium on chromosome 8 (Figure S5).

The strongest association signal was identified for the concentration of cadmium, consisting of 191 significant SNPs that covered a 36.03-Mb interval within a long-range LD region of chromosome 2 (Figure 1A). The peak SNP (S2_157751802; P-value 1.53×10^{-23};

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

162,398,589 bp) for this complex association signal (Table S4), which explained 18% of the phenotypic variance for cadmium, was positioned within the open reading frame (ORF) of a gene (Zm00001d005174) that codes for a protein that belongs to the superfamily of uridine diphosphate-glycosyltransferases (www.maizegdb.org). However, the peak SNP within this recombination suppressed region (Gore *et al.* 2009; Rodgers-Melnick *et al.* 2016) was ~630 kb from the *heavy metal ATPase3* (*hma3*; Zm00001d005190) and *heavy metal ATPase4* (*hma4*; Zm00001d005189) genes (Table S5) that encode proteins with 71% and 66% amino acid sequence identity to OsHMA3, a P_{1B}-type ATPase involved in sequestration of cadmium in root vacuoles of rice (Ueno *et al.* 2010). Additionally, five SNPs significant at 5% FDR within these two genes were associated with cadmium and, on average, in strong LD (mean r^2 of 0.48) with the peak SNP.

To better resolve the expansive association signal resulting from a large-effect locus located in a long range, high LD genomic region, a chromosome-wide multi-locus mixed-model procedure (MLMM) was conducted for cadmium. The resulting optimal model only included the peak SNP (S2_157751802) on chromosome 2 (Table S6). When GWAS was reconducted with this MLMM-selected SNP included as a covariate in the mixed linear model to control for this large-effect locus on chromosome 2, all other previously significant associations on chromosomes 2 and 8 were no longer significant at a genome-wide FDR of 5% (Figure 1B).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

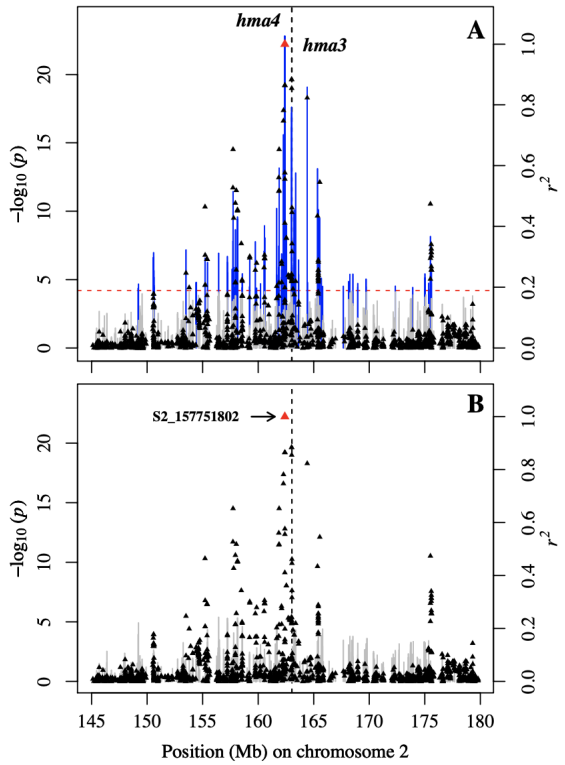

Figure 1. Genome-wide association study for cadmium concentration in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNP) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the r^2 values of each SNP relative to the peak SNP (indicated in red) at 162,398,589 bp (B73 RefGen_v4) on chromosome 2. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical dashed lines indicate the genomic positions of the *heavy metal ATPase 4* (*hma4*; Zm00001d005189; 163,016,710-163,020,248 bp) and *heavy metal ATPase 3* (*hma3*; Zm00001d005190; 163,038,225-163,041,426 bp) genes. These two genes are separated by a physical distance of ~18 kb, thus their positions are not distinguishable at the plotted scale. (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The SNP from the optimal multi-locus mixed-model (S2_157751802) was included as a covariate in the mixed linear model to control for the large-effect locus. None of the tested SNPs were significant at a 5% FDR.

We identified 21 SNPs that spanned a 1.21-Mb region on chromosome 7 that were significantly associated with the concentration of zinc in fresh kernels (Figure 2A). The peak SNP (S7_174515604; *P*-value 3.19×10^{-10}; 180,076,727 bp) for this association signal explained 7% of the phenotypic variance for zinc and was located within the ORF of a gene

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

(Zm00001d022563) that encodes a tetratricopeptide repeat-like superfamily protein (www.maizegdb.org). Notably, this peak SNP was located ~111 kb from the *nicotianamine synthase5* (*nas5*) gene (Zm00001d022557; Table S5), which encodes a class II NAS that presumably contributes to the production of the metal chelator nicotianamine (Zhou *et al.* 2013). Of the 21 detected SNP-zinc associations, SNP S7_174279369, which was located ~160 kb from *nas5* and in moderately strong LD (r^2 = 0.49) with the peak SNP for zinc, also had a near significant association (FDR-adjusted *P*-value 0.06) with the concentration of iron in fresh kernels. When using the chromosome-wide MLMM procedure to better clarify the association signal complex within the 1.21-Mb region on chromosome 7 for zinc, only the peak SNP S7_174515604 was included in the optimal model (Table S6). With the MLMM-selected peak SNP as a covariate, a conditional mixed model analysis did not detect any SNPs significantly associated with zinc (Figure 2B).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

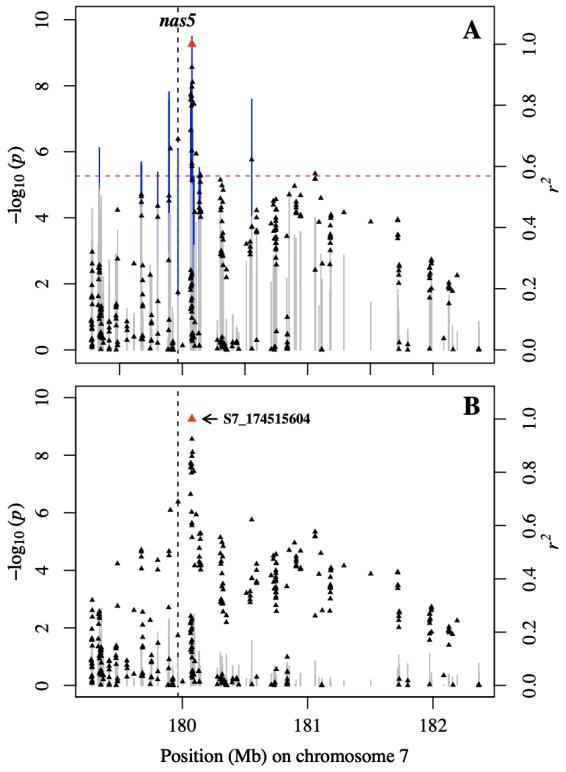

Figure 2. Genome-wide association study for zinc concentration in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNP) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the r^2 values of each SNP relative to the peak SNP (indicated in red) at 180,076,727 bp (B73 RefGen_v4) on chromosome 7. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical dashed line indicates the genomic position of the *nicotianamine synthase 5* (*nas5*; Zm00001d022557;179,964,493-179,965,584 bp) gene. (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The SNP from the optimal multi-locus mixed-model (S7_174515604) was included as a covariate in the mixed linear model to control for the large-effect locus. None of the tested SNPs were significant at a 5% FDR.

Compared to cadmium and zinc, a relatively weaker association signal consisting of seven significant SNPs was identified for boron concentration on chromosome 1. Collectively, these seven SNPs comprised a 98.99-kb interval. The peak SNP (S1_189146031; *P*-value 5.01×10^{-7}; 191,327,920 bp), which was located within the ORF of a gene (Zm00001d031473) encoding a protein with 87% sequence identity to an aminoacylase in rice (Table S5), explained

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

5% of the phenotypic variance for boron concentration. Of the other candidate genes within ± 250 kb of the peak SNP position, a gene (Zm00001d031476) found to be ~45 kb away from the peak SNP encodes a protein with 40-45% amino acid sequence identity to two heavy metal-associated isoprenylated plant proteins (HIPPs) in Arabidopsis (Table S5) that are putative metallochaperones (de Abreu-Neto *et al.* 2013). Indicative of a weaker effect locus, the MLMM procedure at the chromosome-wide level did not select any SNP for the optimal model.

To identify marker-trait associations that may be location-specific, we conducted GWAS for the 15 fresh kernel elemental phenotypes within each location (NY, Figure S6; WI, Figure S7), resulting in significant association signals detected for cadmium (NY and WI), zinc (NY and WI), nickel (NY), and calcium (WI) at 5% FDR (Table S4). In the NY location, the association signal for cadmium consisted of 198 SNPs that defined a 36.89-Mb region on chromosome 2, with the peak SNP (S2_157751802; *P*-value 2.64×10^{-19}; 162,398,589 bp) for the signal the same as detected for cadmium across locations (Table S4). A different peak association signal (S2_159765450; *P*-value 1.98×10^{-15}; 164,415,588 bp) that contained 81 SNPs covering 32.92-Mb on chromosome 2 was identified for cadmium in the WI location (Table S4). The peak SNP for the WI location was ~1.38-Mb from *hma3* and *hma4*, whereas the peak SNP in the NY location was ~630 kb from the same two candidate genes. Within each location, the chromosome-level MLMM procedure was performed for cadmium (Table S6), selecting only the peak SNP that when included as a covariate in a conditional GWAS rendered all other associations on chromosomes 1 (NY), 2 (NY and WI), and 3 (NY and WI) no longer significant.

In both locations, the same SNP served as the peak association signal (NY: *P*-value 4.89×10^{-8}; WI: *P*-value 7.75×10^{-8}) for zinc, which was ~111 kb from *nas5* on chromosome 7 (Table S4). The same SNP was also detected as the peak association signal for zinc across

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

locations. The chromosome-wide MLMM approach selected only the peak SNP in the best model within each location (Table S6). With the peak SNP as a covariate in a conditional GWAS, the other previously detected SNPs for zinc on chromosome 7 in NY (6 SNPs) and WI (3 SNPs) did not remain significant. Similar to the results from conducting the across-location GWAS, SNP S7_174279369, which was ~160 kb from *nas5*, had a weak association (FDR-adjusted *P*-value 0.07) with iron in the NY location. In contrast to the across-location GWAS, however, this SNP was not significantly associated with zinc in either the NY or WI locations (Table S4).

A significant association signal was detected for calcium in the WI location but not the NY location (Table S4). This signal for calcium consisted of four SNPs on chromosome 10. The peak SNP (S10_124069084; *P*-value 6.09×10^{-08}; 125,112,781 bp) was ~36 kb from a gene (Zm00001d025654) that codes for a protein with 47-49% sequence identity to two HIPPs in Arabidopsis (Table S5). Additionally, this peak SNP was selected by the MLMM in the optimal model at the chromosome-wide level (Table S6). In concordance with other conditional GWAS results, no other SNPs remained significant when SNP S10_124069084 was used as a covariate in the mixed linear model.

Of the two field locations, we only detected a significant association for nickel in the NY location. The association signal for nickel on chromosome 9 consisted of four SNPs, with the peak SNP (S9_2213924; *P*-value 2.80×10^{-08}; 1,934,330 bp) for the signal selected by the MLMM in the optimal model (Tables S4 and S6). The peak SNP was within the ORF of a gene (Zm00001d044768) that encodes a protein 39-47% identical at the amino acid sequence level to three members of the NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER (NRT1/PTR) family (NPF) in Arabidopsis (Table S5) that transport nitrate, amino acids, and hormones (Léran

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

et al. 2014). Additionally, this peak SNP was ~44 kb from a gene (Zm00001d044771) coding a protein with 49-52% sequence identity to three matrix metalloproteinases (MMPs) in Arabidopsis (Table S5) that have a zinc-binding sequence (Marino and Funk 2012). All other significant SNPs on chromosome 9 and a single significant SNP on chromosome 5 were no longer significantly associated with nickel in a conditional mixed linear model with the peak SNP as a covariate.

Whole-genome prediction

We evaluated the predictive ability of WGP with 147,762 SNP markers for the across-location concentrations of 15 elements that had been scored on fresh kernels of the 401 inbred lines. The 15 elements had an average predictive ability of 0.37, ranging in abilities from 0.19 for rubidium to 0.52 for copper (Table 3). The predictive abilities were above average for iron (0.45) and zinc (0.49), suggesting that genomic selection could be used to increase the concentration of both nutritionally limiting microelements in fresh sweet corn kernels. A strong positive Pearson's correlation coefficient ($r = 0.62$ P-value < 0.05) was found between heritability estimates and predictive abilities for the 15 elemental phenotypes. Given the detection of significant differences among endosperm mutation types for 10 of the 15 elemental phenotypes (Table 2), endosperm mutation type was tested as an included covariate in WGP models, but changes to prediction abilities from its inclusion were zero to negligible (Table 3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Table 3. Mean predictive abilities (standard deviation) of whole-genome prediction models for 15 fresh kernel elemental phenotypes in the sweet corn association panel.

Trait	GBLUP	GBLUP with endosperm mutation type covariate
Boron	0.23 (0.02)	0.22 (0.02)
Cadmium	0.41 (0.02)	0.43 (0.02)
Calcium	0.27 (0.03)	0.27 (0.03)
Copper	0.52 (0.01)	0.53 (0.01)
Iron	0.45 (0.02)	0.45 (0.02)
Magnesium	0.40 (0.02)	0.40 (0.02)
Manganese	0.48 (0.02)	0.47 (0.02)
Molybdenum	0.25 (0.03)	0.25 (0.03)
Nickel	0.40 (0.02)	0.40 (0.02)
Phosphorus	0.47 (0.02)	0.47 (0.02)
Potassium	0.42 (0.02)	0.42 (0.02)
Rubidium	0.19 (0.03)	0.18 (0.03)
Strontium	0.32 (0.02)	0.32 (0.02)
Sulfur	0.26 (0.02)	0.27 (0.02)
Zinc	0.49 (0.02)	0.49 (0.02)
Average	0.37	0.37

Discussion

Maintaining elemental homeostasis is critical for plants to realize optimal growth and complete their life cycle (Marschner 2012). Additionally, the elemental content and composition of edible plant parts are influenced by genetic and environmental factors (Watanabe *et al.* 2007; Baxter and Dilkes 2012; Baxter *et al.* 2014). Several genes responsible for natural variation of elemental levels in root and shoot tissues have been identified and characterized in plants (Huang and Salt 2016; Yang *et al.* 2018), but considerable effort remains to pinpoint the genes regulating elemental levels in seed of crops. To further this research, we examined the extent of phenotypic

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

variation for elemental concentrations in fresh kernels and performed a GWAS to identify candidate genes controlling this phenotypic variability in a sweet corn association panel. We also evaluated the ability of genome-wide markers to predict elemental concentrations, providing insights into the potential of genomic selection for optimizing the elemental profile of fresh kernels for human health and nutrition, especially iron and zinc. To the best of our knowledge, this work is the most extensive quantitative genetic analysis of elemental concentrations in fresh kernels of sweet corn.

The rank order of average concentrations for the measured elements in fresh kernels from the sweet corn association panel was highly concordant with that observed for the same elements in physiologically mature grain from the maize Ames panel of non-sweet corn tropical and temperate inbred lines (Wu *et al.* 2021 in review) and the B73 (dent; *Su1*) x IL14H (sweet corn; *su1*) recombinant inbred line (RIL) family of the U.S. NAM panel (Baxter *et al.* 2014). Our sweet corn association panel showed a range of 13.66-28.11 and 18.58-32.10 μg g^{-1} on a dry weight basis for iron and zinc, respectively. However, these ranges were both lower and narrower than that found for iron (14.62 - 36.33 μg g^{-1}; 3.29 S.D.) and zinc (12.59 - 52.32 μg g^{-1}; 4.36 S.D.) in physiologically mature grain from the maize Ames panel (Wu *et al.* 2021 in review). With consideration of losses from processing and bioavailability, the recommended target level iron and zinc grain concentrations established by HarvestPlus are 60 and 38 μg g^{-1} dry weight, respectively, for developing biofortified maize based on achieving ~30-40% of estimated average requirements for adult women (nonpregnant, nonlactating) when consuming 400 g d^{-1} of whole maize grain (Bouis and Welch 2010). Although the maximum iron concentration observed in our sweet corn association panel (28.11 μg g^{-1} dry weight) is ~2-fold lower than the HarvestPlus breeding target, there are six inbred lines that have zinc

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

concentrations ranging from 30.00 to 32.10 μg g^{-1} dry weight. As it relates to the consumption of fresh sweet corn, the estimated maximum fresh kernel concentrations observed in our association panel would provide approximately 4-9% and 7-10% of the recommended daily allowance of iron and zinc, respectively, for adult non-elderly women (nonpregnant, nonlactating) and men when consuming 100 g (~75% water) of uncooked fresh sweet corn.

We assessed whether the concentrations of elements differed significantly among endosperm mutation group types, given that certain elements were not distributed equally amongst the embryo and endosperm of maize kernels (Reviewed in Suri and Tanumihardjo 2016). Of the 15 elements, copper, iron, manganese, potassium, and sulfur were highly significantly different ($P < 0.001$) between two or more endosperm mutation type groups (*su1*, *sh2*, and *su1sh2*) (Table 2). Relatedly, Baxter *et al.* (2014) showed that the content for four (iron, manganese, potassium, and sulfur) of these five elements significantly differed ($P < 0.0005$) between visibly 'shrunken' (*su1*/*su1*) and 'non-shrunken' (*Su1*/*su1* or *Su1*/*Su1*) kernels harvested at physiological maturity. As it relates to the spatial distribution of elements in sweet corn kernels, Cheah *et al.* (2019) analyzed immature (21 DAP) kernels from a single sweet corn (*sh2*) variety via synchrotron-based X-ray fluorescence microscopy to reveal that potassium and calcium were generally present throughout the kernel, sulfur concentrated mainly in the axis of the embryo and the periphery of the endosperm, and the scutellum of the embryo had at least 20-fold higher concentrations of phosphorus, iron, zinc, and manganese than in the endosperm. Notably, Cheah *et al.* (2019) also showed that these spatial distribution maps for elements were highly similar to those of immature maize (non-sweet corn) kernels. Given their findings, the observed significant differences in elemental concentrations among endosperm mutation type

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

groups in our study were more likely due to genetic variability in accumulation of elements rather than the differential spatial distribution of elements in fresh kernels.

Conducting GWAS across locations for the concentrations of elements in fresh kernels of the sweet corn association panel resulted in the identification of candidate genes associated with cadmium and zinc at the genome-wide level. Of these elements, the strongest association signal was for cadmium, having an association signal on chromosome 2 that spread more than 35-Mb across a recombinationally inert genomic region. Additionally, the peak SNP of this association signal co-localized with the single across-location QTL detected for grain cadmium concentration in the maize NAM panel (Tables S4 and S7), but the GWAS resolution for this region in the NAM panel (Table S8) was too limiting to convincingly identify an underlying causal gene (Ziegler *et al.* 2017). In our sweet corn association panel, however, two likely candidate causal genes were identified, *hma3* and *hma4*, both having SNPs in moderately strong LD with the peak SNP of the association signal. These genes are two of a 12-member gene family encoding HMAs in the genome of maize inbred line B73 (Cao *et al.* 2019).

Of the 12 HMA genes, the proteins encoded by *hma3* and *hma4* have high sequence identity (Table S5) to the P_{1B}-type ATPase, OsHMA3—a tonoplast-localized zinc/cadmium transporter that has been shown to be expressed in rice roots, mediates cadmium and zinc vacuolar sequestration and, as such, participates in zinc homeostasis and root-to-shoot cadmium translocation (Ueno *et al.* 2010; Miyadate *et al.* 2011; Sasaki *et al.* 2014; Cai *et al.* 2019). The loss-of-function of *OsHMA3* has been associated with cadmium accumulation in rice grains, whereas low-cadmium rice cultivars express a functional *OsHMA3* (Ueno *et al.* 2010). Furthermore, Ueno *et al.* (2010) have shown that the overexpression of *OsHMA3* selectively decreased the accumulation of cadmium, but not other elements in the grain. In a follow-up

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

study, Sasaki *et al.* (2014) showed that overexpression of *OsHMA3* resulted in sequestration of both cadmium and zinc in rice root vacuoles, but the concentration of zinc in shoots was unaffected through the constitutive upregulation of transporter genes having putative involvement in the uptake and translocation of zinc. In agreement with Sasaki *et al.* (2014), we found no evidence of the large-effect locus spanning *hma3* and *hma4* having a significant association with zinc concentration in fresh sweet corn kernels. Importantly, Cao *et al.* (2019) identified several polymorphisms within *hma3* to be significantly associated with leaf cadmium concentration at the seedling and adult plant stages in a maize diversity panel. Moreover, they further showed the expression level of *hma3* to be highly upregulated in the roots of B73 in response to cadmium stress, whereas the expression of *hma4* was undetectable in roots under the same conditions (Cao *et al.* 2019). In light of this, we propose that *hma3* is the more likely of the two genes to have played a key genetic role in the accumulation of cadmium in fresh kernels.

The *nas5* gene was found to be within ± 250 kb of the peak SNP for the across-location zinc and iron association signals on chromosome 7. These findings cohere with the GWAS results of Ziegler *et al.* (2017) and Wu *et al.* (2021 in review) from the U.S. maize NAM (Tables S4, S7, and S8) and Ames mapping panels, respectively, that implicated *nas5* as a possible pleiotropic controller for the concentrations of both zinc and iron in physiologically mature grain samples. As one of nine gene family members in the B73 reference genome, *nas5* phylogenetically groups together with *nas3* and *nas4*, which together comprise class II *nas* genes (Zhou *et al.* 2013). The NAS enzyme encoded by *nas5* is hypothesized to be involved in the production of the non-proteinogenic amino acid, nicotianamine, an efficient chelator of transition metals including zinc and iron (Takahashi *et al.* 2003; Curie *et al.* 2009; Swamy *et al.* 2016). In addition to a suggested role in intracellular metal homeostasis, nicotianamine facilitates

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

phloem-based metal delivery from source (e.g., leaves) to sink (e.g., seeds) tissues (Takahashi *et al.* 2003; Curie *et al.* 2009; Swamy *et al.* 2016). Nicotianamine also serves as a precursor for the synthesis of root-exuded mugineic acid-type phytosiderophores that chelate divalent metals for eventual root uptake (Curie *et al.* 2009; Swamy *et al.* 2016).

Consistent with a proposed role of *nas5* in long-distance metal transport rather than uptake into roots, Zhou *et al.* (2013) showed that transcripts of class II *nas* genes including *nas5* accumulated mainly in maize leaves and sheath, whereas class I *nas* genes were predominantly expressed in maize roots. Interestingly, of the three class II *nas* genes, *nas5* was more highly expressed in maize stems, further suggesting its contribution to long-distance metal transport and perhaps its contribution to metal loading to seeds (Zhou *et al.* 2013). In addition, the transcriptional expression level of *nas5* was downregulated by iron deficiency in both shoots and roots, but upregulated under excess iron and zinc in roots. Activation tagging of *OsNAS3*, the closest rice homolog of *nas5* (Zhou *et al.* 2013), produced an increased level of nicotianamine that resulted in elevated levels of zinc and iron in shoots, roots, and seeds of activation-tagged rice plants (Lee *et al.* 2009). Despite the lack of functional validation, results from the maize association mapping and transgenic rice studies strongly support the nomination of *nas5* as a causal gene for controlling zinc and iron concentrations in fresh sweet corn kernels.

Similar to our findings with *nas5* in the sweet corn association panel, Wu *et al.* (2021 in review) also identified a stronger association signal of *nas5* with zinc relative to iron in the maize Ames panel. This finding is not entirely surprising considering that the affinity constants (Kd) of complexes of nicotianamine with zinc is higher than for nicotianamine with iron (Curie *et al.* 2009; Gayomba *et al.* 2015). Therefore, it is plausible to propose that equimolar concentrations of iron and zinc would result in the selection of zinc for nicotianamine over iron (Curie *et al.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

2009; Gayomba *et al.* 2015). It is also noteworthy that the concentration of zinc in the phloem sap is thought to be higher than the concentration of iron (Reviewed in Gayomba *et al.* 2015), thus reinforcing the suggested role of nicotianamine and *nas5* in zinc accumulation in fresh kernels.

In contrast to the highly probable causality of the *hma3* and *nas5* loci detected via GWAS across and within both locations, there is only moderately compelling evidence for the genetic involvement of identified candidate genes for the concentrations of boron (across locations), calcium (WI), and nickel (NY). Of the genes within ± 250 kb of the peak SNP for the boron association signal on chromosome 1, two genes encoding a putative protein with sequence identity to either an aminoacylase (Zm00001d031473) or HIPP (Zm00001d031476) were the most plausible candidates. Under boron deficiency, aminoacylases (metalloenzymes involved in amino acid metabolism) have been shown to have decreased protein and increased transcript levels in *Brassica napus* L. roots and *Citrus sinensis* leaves, respectively (Wang *et al.* 2010; Lu *et al.* 2015). Transcriptome profiling revealed a *HIPP* to be upregulated in leaves and roots of black poplar (*Populus nigra* L.) grown under boron toxicity (Yıldırım and Uylaş 2016), which perhaps is not surprising given that HIPPs are metallochaperones involved in the transport of metallic ions and response to abiotic stresses (de Abreu-Neto *et al.* 2013). Despite these findings in other plant systems, the exact mechanism by which the identified aminoacylase and HIPP would have contributed to the accumulation of boron in fresh kernels is unknown but nevertheless merits further experimental investigation.

Although the peak SNP associated with calcium concentration in the WI location resided within a genomic region that lacked a definitive candidate gene, it was located ~36 kb from a gene (Zm00001d025654) encoding a putative HIPP. However, to our knowledge, HIPPs have

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

never been experimentally shown to bind Ca^{2+} (de Abreu-Neto *et al.* 2013), thus an *in vitro* study would be needed to determine whether Ca^{2+} is bound by the putative HIPP that Zm00001d025654 encodes. It is interesting that collectively, two different HIPP candidate genes were identified for the concentrations of boron and calcium, but these proteins would not be expected to have similar roles given that calcium and boron have different chemical and physiological properties (Marschner 2012). Regardless, these findings open up new avenues of inquiry that could deepen our understanding of the genetic basis of boron and calcium accumulation in fresh sweet corn kernels.

In contrast to calcium, the nickel association signal for the NY location coincided with association signals detected for nickel grain concentrations in the maize NAM (Tables S4, S7, and S8) and Ames panels (Ziegler *et al.* 2017; Wu *et al.* 2021 in review). This still genetically-unresolved signal consisted of two possible candidate genes, an MMP (Zm00001d044771) and NPF member (Zm00001d044768). Even though MMPs could conceivably bind nickel in place of zinc (Cerdà-Costa and Gomis-Rüth 2014), their proteolytic activities to remodel the extracellular matrix (Marino and Funk 2012) lack a clear connection to nickel transport or accumulation. Additionally, the putative NPF member encoded by Zm00001d044768 is hypothesized to transport nitrate given its high sequence identity to members of the NPF5 subfamily in Arabidopsis (Table S5) (Niño-González *et al.* 2019), thus reducing but not completely eliminating the possibility that this yet-to-be characterized protein transports nickel or a substrate that binds nickel.

We showed that, on average, moderate predictive abilities were achieved through the application of WGP for the across-location concentrations of the 15 elements in fresh sweet corn kernels. Additionally, these predictive abilities were found to be strongly correlated with

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

heritability estimates, which coheres with expectations (Combs and Bernardo 2013). In accordance with these results, Wu *et al.* (2021 in review) analyzed 11 of these 15 elements in mature grain samples from the maize Ames association panel that had been evaluated in a single location across two years, finding that the moderate predictive abilities of the 11 elements from WGP had a strong correlation with their heritabilities. Also, the prediction abilities presented in Table 3 suggest that endosperm mutation type did not need to be considered as a covariate in WGP models to capture the genetic differences between the three groups at the genome-wide marker density employed in this study. Given that we observed significant G×E interaction for boron, copper, iron, magnesium, and rubidium, accounting for G×E in WGP models could result in slightly improved predictive abilities for these five elements. In support of this supposition, a multi-environment model incorporating G×E resulted in higher average prediction abilities for the concentration of zinc in kernels from a tropical maize inbred panel and a double haploid population compared to those from single-environment models (Mageto *et al.* 2020). Therefore, this should be an area of further exploration when conducting multi-environment genomic selection for elemental phenotypes whether at the immature or mature stages of kernel development.

The results from GWAS can be used to better inform how best to successfully implement genomic selection for the concentrations of elements in fresh sweet corn kernels. With the exception of cadmium (18%) and zinc (7%), which each had a single locus explaining more than 5% of the phenotypic variance in the across-location GWAS, we observed relatively weaker association signals for boron, iron, and the other 11 elements in fresh kernels. Furthermore, we did not identify all of the strong association signals for the grain concentrations of boron, copper, manganese, molybdenum, nickel, and zinc that had been detected in the maize Ames panel (Wu

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

et al. 2021 in review). Compared to the Ames panel study of Wu *et al.* 2021 (in review) that had more than 2,000 maize inbred lines, it is likely that the size of the sweet corn association panel in our study was too underpowered to identify these particular loci, whether as a result of their lower allele frequencies and/or weaker effects. Regardless, we posit that these elemental phenotypes are generally more polygenic than carotenoid and tocochromanol levels in fresh kernels of sweet corn that have a more oligogenic inheritance (Baseggio *et al.* 2019, 2020). Therefore, genomic selection is more advisable than marker-assisted selection as a breeding approach for selecting for the concentration of elements in fresh kernels (Lorenz *et al.* 2011; Desta and Ortiz 2014). However, it is still worthwhile to assess the inclusion of large-effect loci as fixed effects such as those for cadmium and zinc in WGP models, as it could result in higher prediction abilities in specific sweet corn breeding populations (Bernardo 2014).

Conclusions

We used a sweet corn association panel to study the quantitative genetics of natural variation for the concentrations of 15 elements in fresh kernels. Through an across-location GWAS, we strongly implicated the candidate causal genes *nas5* with iron/zinc and *hma3* with cadmium. Given that iron and zinc accumulation in fresh kernels have a partially shared genetic basis, the genetic correlation between these two phenotypes can be leveraged with multi-trait genomic selection approaches to possibly exceed the prediction accuracy of single-trait genomic selection (Jia and Jannink 2012) for simultaneous genetic gains in zinc and iron concentrations. Such efforts would help to address iron and zinc deficiencies of women, children, and older adults in the U.S. (Clark 2008) where sweet corn is highly consumed as a fresh vegetable. Importantly, the across- and within-location association signals at the *nas5* and *hma3* loci were specific to zinc/iron and cadmium, respectively. This suggests that genomic selection for lower cadmium

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

accumulation to reduce possible toxicity should not influence zinc accumulation in the kernel. Even though 100 g of fresh sweet corn (medium-sized ear) with the maximum concentration of cadmium found in this panel is estimated to provide less than 2% of the provisional tolerable intake for this element in a day (0.8 µg/kg bw/day) (JECFA 2011) when consumed by a 70 kg person, efforts should be dedicated towards breeding sweet corn that has lower cadmium but higher zinc and iron, considering that sweet corn can be grown in regions with naturally elevated cadmium levels, or with low zinc and iron levels.

Author Contributions

M.B., O.K.V., and M.A.G. co-wrote the manuscript; M.B. led the data analysis; G.Z. performed the ICP-MS analyses and elemental quantifications; N.K. and M.M. provided overall management of panel growth (planting, pollination, harvesting); M.B., M.M., and J.C. generated the marker datasets; I.B. oversaw the ICP-MS analyses and provided biological interpretation; C.R.B. and J.P.H. uplifted the NAM JL-GWAS results; D.W. integrated and provided biological interpretation of the NAM JL-GWAS results; O.K.V. provided biological interpretation of the candidate gene associations; W.F.T., M.M., and E.S.B. constructed the association panel; E.S.B., M.E.S., W.F.T., and M.A.G. conceived and designed the project; M.A.G. oversaw the data analysis, project management, design, and coordination.

Acknowledgements

This research was supported by the National Institute of Food and Agriculture (NIFA); the USDA Hatch under accession numbers 100397 (M.A.G.), 1010428 (M.A.G.), 1013637 (M.A.G.), 1013641 (M.A.G.), and 142 AAC6861 072600 4 (W.F.T.), HarvestPlus (M.A.G.), the National Science Foundation (IOS-1546657 to C.R.B. and M.A.G.), USDA-NIFA grant numbers 2021-67013-33798 (O.K.V.) and 2018-67013-27418 (O.K.V.), Cornell University startup funds

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

(M.A.G.), and the USDA-ARS (E.S.B. and I.B.). We are grateful to Jenna Hershberger for providing valuable feedback on an earlier version of the manuscript. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. We thank the current and past members of the Tracy and Gore labs for their efforts in pollination, harvest, and sample preparation. We also thank Lily Hislop and Patrick Flannery for sharing information from the Wisconsin field trials.

Conflict of Interest

The authors declare no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

References

de Abreu-Neto, J. B., A. C. Turchetto-Zolet, L. F. V. de Oliveira, M. H. B. Zanettini, and M. Margis-Pinheiro, 2013 Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. FEBS J. 280: 1604–1616.

Akdemir, D., and U. G. Okeke, 2015 EMMREML: Fitting mixed models with known covariance structures. Vienna, Austria: R Core Development Team. https://CRAN.R-project.org/package=EMMREML.

Axelsen, K. B., and M. G. Palmgren, 2001 Inventory of the superfamily of P-type ion pumps in Arabidopsis. Plant Physiol. 126: 696–706.

Baseggio, M., M. Murray, M. Magallanes-Lundback, N. Kaczmar, J. Chamness *et al.*, 2019 Genome-wide association and genomic prediction models of tocochromanols in fresh sweet corn kernels. Plant Genome 12: 180038.

Baseggio, M., M. Murray, M. Magallanes-Lundback, N. Kaczmar, J. Chamness *et al.*, 2020 Natural variation for carotenoids in fresh kernels is controlled by uncommon variants in sweet corn. Plant Genome 13: e20008.

Baxter, I., 2009 Ionomics: studying the social network of mineral nutrients. Curr. Opin. Plant Biol. 12: 381–386.

Baxter, I., 2015 Should we treat the ionome as a combination of individual elements, or should we be deriving novel combined traits? J. Exp. Bot. 66: 2127–2131.

Baxter, I., and B. P. Dilkes, 2012 Elemental profiles reflect plant adaptations to the environment. Science 336: 1661–1663.

Baxter, I. R., G. Ziegler, B. Lahner, M. V. Mickelbart, R. Foley *et al.*, 2014 Single-kernel ionomic profiles are highly heritable indicators of genetic and environmental influences on

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

elemental accumulation in maize grain (*Zea mays*). PLoS ONE 9: e87628.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 57: 289–300.

Bernardo, R., 2014 Genomewide selection when major genes are known. Crop Sci. 54: 68–75.

Bhullar, N. K., and W. Gruissem, 2013 Nutritional enhancement of rice for human health: the contribution of biotechnology. Biotechnol. Adv. 31: 50–57.

Bouis, H. E., and R. M. Welch, 2010 Biofortification—A sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. Crop Sci. 50: S–20–S–32.

Box, G. E. P., and D. R. Cox, 1964 An analysis of transformations. J. R. Stat. Soc. Series B Stat. Methodol. 26: 211–243.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss *et al.*, 2007 TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633–2635.

Broadley, M., P. Brown, I. Cakmak, Z. Rengel, and F. Zhao, 2012 Function of Nutrients, pp. 191–248 in *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*, Academic Press, San Diego.

Cai, H., S. Huang, J. Che, N. Yamaji, and J. F. Ma, 2019 The tonoplast-localized transporter OsHMA3 plays an important role in maintaining Zn homeostasis in rice. J. Exp. Bot. 70: 2717–2725.

Cao, Y., X. Zhao, Y. Liu, Y. Wang, W. Wu *et al.*, 2019 Genome-wide identification of *ZmHMAs* and association of natural variation in *ZmHMA2* and *ZmHMA3* with leaf cadmium accumulation in maize. PeerJ 7: e7877.

Cerdà-Costa, N., and F. X. Gomis-Rüth, 2014 Architecture and function of metallopeptidase

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

catalytic domains. Protein Sci. 23: 123–144.

Chan-Rodriguez, D., and E. L. Walker, 2018 Analysis of yellow striped mutants of *Zea mays* reveals novel loci contributing to iron deficiency chlorosis. Front. Plant Sci. 9: 157.

Cheah, Z. X., P. M. Kopittke, S. M. Harper, T. J. O'Hare, P. Wang *et al.*, 2019 In situ analyses of inorganic nutrient distribution in sweetcorn and maize kernels using synchrotron-based X-ray fluorescence microscopy. Ann. Bot. 123: 543–556.

Chen, J., and Z. Chen, 2008 Extended Bayesian information criteria for model selection with large model spaces. Biometrika 95: 759–771.

Clark, S. F., 2008 Iron deficiency anemia. Nutr. Clin. Pract. 23: 128–141.

Combs, E., and R. Bernardo, 2013 Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. Plant Genome 6: 1–7.

Curie, C., G. Cassin, D. Couch, F. Divol, K. Higuchi *et al.*, 2009 Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. Ann. Bot. 103: 1–11.

Davies, L., and U. Gather, 1993 The identification of multiple outliers. J. Am. Stat. Assoc. 88: 782–792.

Desta, Z. A., and R. Ortiz, 2014 Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci. 19: 592–601.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6: e19379.

Gayomba, S. R., Z. Zhai, H.-I. Jung, and O. K. Vatamaniuk, 2015 Local and systemic signaling of iron status and its interactions with homeostasis of other essential elements. Front. Plant

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Sci. 6: 716.

Gilmour, A. R., B. J. Gogel, B. R. Cullis, R. Thompson, D. Butler *et al.*, 2009 ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9: e90346.

Gore, M. A., J.-M. Chia, R. J. Elshire, Q. Sun, E. S. Ersoz *et al.*, 2009 A first-generation haplotype map of maize. Science 326: 1115–1117.

Guo, R., T. Dhliwayo, E. K. Mageto, N. Palacios-Rojas, M. Lee *et al.*, 2020 Genomic prediction of kernel zinc concentration in multiple maize populations using genotyping-by-sequencing and repeat amplification sequencing markers. Front. Plant Sci. 11: 534.

Hawkesford, M., W. Horst, T. Kichey, H. Lambers, J. Schjoerring *et al.*, 2012 Functions of macronutrients, pp. 135–189 in *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*, Academic Press, San Diego.

Higuchi, K., K. Suzuki, H. Nakanishi, H. Yamaguchi, N. K. Nishizawa *et al.*, 1999 Cloning of nicotianamine synthase genes, novel genes involved in the biosynthesis of phytosiderophores. Plant Physiol. 119: 471–480.

Hill, W. G., and B. S. Weir, 1988 Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33: 54–78.

Hindu, V., N. Palacios-Rojas, R. Babu, W. B. Suwarno, Z. Rashid *et al.*, 2018 Identification and validation of genomic regions influencing kernel zinc and iron in maize. Theor. Appl. Genet. 131: 1443–1457.

Hirschi, K. D., 2020 Genetically modified plants: Nutritious, sustainable, yet underrated. J. Nutr. 150: 2628–2634.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Holland, J. B., W. E. Nyquist, and C. T. Cervantes-Martínez, 2003 Estimating and interpreting heritability for plant breeding: an update. Plant Breed. Rev. 22: 9–112.

Huang, X.-Y., and D. E. Salt, 2016 Plant ionomics: From elemental profiling to environmental adaptation. Mol. Plant 9: 787–797.

Hung, H.-Y., C. Browne, K. Guill, N. Coles, M. Eller *et al.*, 2012 The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. Heredity 108: 490–499.

Institute of Medicine, 2001 *Dietary Reference Intakes for Vitamin A, Vitamin K, Arsenic, Boron, Chromium, Copper, Iodine, Iron, Manganese, Molybdenum, Nickel, Silicon, Vanadium, and Zinc*. The National Academies Press, Washington, DC.

JECFA, 2011 Evaluations of the joint FAO/WHO expert committee on food additives. 73th Report of the Joint FAO/WHO Expert Committee on Food Additive. WHO Technical Report Series 960.

Jia, Y., and J.-L. Jannink, 2012 Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics 192: 1513–1522.

Kenward, M. G., and J. H. Roger, 1997 Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53: 983–997.

Khan, M. A., N. Castro-Guerrero, and D. G. Mendoza-Cozatl, 2014 Moving toward a precise nutrition: preferential loading of seeds with essential nutrients over non-essential toxic elements. Front. Plant Sci. 5: 51.

Klatte, M., M. Schuler, M. Wirtz, C. Fink-Straube, R. Hell *et al.*, 2009 The analysis of Arabidopsis nicotianamine synthase mutants reveals functions for nicotianamine in seed iron loading and iron deficiency responses. Plant Physiol. 150: 257–271.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Kurtz, S., 2003 The Vmatch large scale sequence analysis software - a manual.

Lee, S., U. S. Jeon, S. J. Lee, Y.-K. Kim, D. P. Persson *et al.*, 2009 Iron fortification of rice seeds through activation of the nicotianamine synthase gene. Proc. Natl. Acad. Sci. U. S. A. 106: 22014–22019.

Léran, S., K. Varala, J.-C. Boyer, M. Chiurazzi, N. Crawford *et al.*, 2014 A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. Trends Plant Sci. 19: 5–9.

Li, S., X. Liu, X. Zhou, Y. Li, W. Yang *et al.*, 2019 Improving zinc and iron accumulation in maize grains using the zinc and iron transporter ZmZIP5. Plant Cell Physiol. 60: 2077–2085.

Linus Pauling Institute, 2016 Micronutrient Information Center: Iron. https://lpi.oregonstate.edu/mic/minerals/iron. Accessed February 5, 2021.

Linus Pauling Institute, 2019 Micronutrient Information Center: Zinc. https://lpi.oregonstate.edu/mic/minerals/zinc. Accessed February 5, 2021.

Lipka, A. E., M. A. Gore, M. Magallanes-Lundback, A. Mesberg, H. Lin *et al.*, 2013 Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 3: 1287–1299.

Lipka, A. E., F. Tian, Q. Wang, J. Peiffer, M. Li *et al.*, 2012 GAPIT: genome association and prediction integrated tool. Bioinformatics 28: 2397–2399.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberger, 2006 Appendix 1: Linear mixed model theory. SAS for mixed models. SAS Institute Inc., Cary, NC 733–756.

Lombi, E., K. G. Scheckel, J. Pallon, A. M. Carey, Y. G. Zhu *et al.*, 2009 Speciation and distribution of arsenic and localization of nutrients in rice grains. New Phytol. 184: 193–

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

201.

Lombi, E., E. Smith, T. H. Hansen, D. Paterson, M. D. de Jonge *et al.*, 2011 Megapixel imaging of (micro)nutrients in mature barley grains. J. Exp. Bot. 62: 273–282.

Lönnerdal, B., 2000 Dietary factors influencing zinc absorption. J. Nutr. 130: 1378S–1383S.

Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi *et al.*, 2011 Genomic selection in plant breeding: knowledge and prospects. Adv. Agron. 110: 77–123.

Lu, Y.-B., Y.-P. Qi, L.-T. Yang, J. Lee, P. Guo *et al.*, 2015 Long-term boron-deficiency-responsive genes revealed by cDNA-AFLP differ between *Citrus sinensis* roots and leaves. Front. Plant Sci. 6: 585.

Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA.

Maathuis, F. J. M., 2009 Physiological functions of mineral macronutrients. Curr. Opin. Plant Biol. 12: 250–258.

Mageto, E. K., J. Crossa, P. Pérez-Rodríguez, T. Dhliwayo, N. Palacios-Rojas *et al.*, 2020 Genomic prediction with genotype by environment interaction analysis for kernel zinc concentration in tropical maize germplasm. G3 10: 2629–2639.

Marino, G., and C. Funk, 2012 Matrix metalloproteinases in plants: a brief overview. Physiol. Plant. 145: 196–202.

Marschner, P., 2012 *Marschner's Mineral Nutrition of Higher Plants (Third Edition)*. Academic Press, San Diego.

Mengel, K., and E. A. Kirkby, 2012 *Principles of Plant Nutrition*. Springer Science & Business Media.

Miller, J. L., 2013 Iron deficiency anemia: a common and curable disease. Cold Spring Harb.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Perspect. Med. 3: a011866.

Miyadate, H., S. Adachi, A. Hiraizumi, K. Tezuka, N. Nakazawa *et al.*, 2011 OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles. New Phytologist 189: 190–199.

Mocchegiani, E., J. Romeo, M. Malavolta, L. Costarelli, R. Giacconi *et al.*, 2013 Zinc: dietary intake and impact of supplementation on immune function in elderly. Age 35: 839–860.

Murgia, I., P. Arosio, D. Tarantino, and C. Soave, 2012 Biofortification for combating "hidden hunger" for iron. Trends Plant Sci. 17: 47–55.

Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, 1996 *Applied linear statistical models*. Irwin Chicago.

Niño-González, M., E. Novo-Uzal, D. N. Richardson, P. M. Barros, and P. Duque, 2019 More transporters, more substrates: The *Arabidopsis* major facilitator superfamily revisited. Mol. Plant 12: 1182–1202.

Pongrac, P., K. Vogel-Mikuš, L. Jeromel, P. Vavpetič, P. Pelicon *et al.*, 2013 Spatially resolved distributions of the mineral elements in the grain of tartary buckwheat (*Fagopyrum tataricum*). Food Res. Int. 54: 125–131.

Prasad, A. S., 2014 Impact of the discovery of human zinc deficiency on health. J. Trace Elem. Med. Biol. 28: 357–363.

R Core Team, 2019 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodgers-Melnick, E., D. L. Vera, H. W. Bass, and E. S. Buckler, 2016 Open chromatin reveals the functional maize genome. Proc. Natl. Acad. Sci. U. S. A. 113: E3177–E3184.

Romay, M. C., M. J. Millard, J. C. Glaubitz, J. A. Peiffer, K. L. Swarts *et al.*, 2013

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14: R55.

Sandstead, H. H., and J. C. Smith Jr, 1996 Deliberations and evaluations of approaches, endpoints and paradigms for determining zinc dietary recommendations. J. Nutr. 126: 2410S–2418S.

Sasaki, A., N. Yamaji, and J. F. Ma, 2014 Overexpression of *OsHMA3* enhances Cd tolerance and expression of Zn transporter genes in rice. J. Exp. Bot. 65: 6013–6021.

Schroeder, J. I., E. Delhaize, W. B. Frommer, M. L. Guerinot, M. J. Harrison *et al.*, 2013 Using membrane transporters to improve crops for sustainable food production. Nature 497: 60–66.

Schwarz, G., 1978 Estimating the dimension of a model. Ann. Stat. 6: 461–464.

Segura, V., B. J. Vilhjálmsson, A. Platt, A. Korte, Ü. Seren *et al.*, 2012 An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat. Genet. 44: 825–830.

Singh, S., P. Parihar, R. Singh, V. P. Singh, and S. M. Prasad, 2016 Heavy metal tolerance in plants: Role of transcriptomics, proteomics, metabolomics, and ionomics. Front. Plant Sci. 6: 1143.

Sun, G., C. Zhu, M. H. Kramer, S.-S. Yang, W. Song *et al.*, 2010 Variation explained in mixed-model association mapping. Heredity 105: 333–340.

Suri, D. J., and S. A. Tanumihardjo, 2016 Effects of different processing methods on the micronutrient and phytochemical contents of maize: From A to Z. Compr. Rev. Food Sci. Food Saf. 15: 912–926.

Swamy, B. P. M., M. A. Rahman, M. A. Inabangan-Asilo, A. Amparado, C. Manito *et al.*, 2016

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Advances in breeding for high grain zinc in rice. Rice 9: 49.

Swarts, K., H. Li, J. A. Romero Navarro, D. An, M. C. Romay *et al.*, 2014 Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7: 2014–2005.

Takahashi, M., Y. Terada, I. Nakai, H. Nakanishi, E. Yoshimura *et al.*, 2003 Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. Plant Cell 15: 1263–1280.

Ueno, D., N. Yamaji, I. Kono, C. F. Huang, T. Ando *et al.*, 2010 Gene limiting cadmium accumulation in rice. Proc. Natl. Acad. Sci. U. S. A. 107: 16500–16505.

USDA-ARS, 2019 FoodData Central. https://fdc.nal.usda.gov. Accessed Feb 04, 2021.

USDA-NASS, 2018 Vegetables 2017 summary. https://usda.library.cornell.edu/concern/publications/02870v86p?locale=en. Accessed Feb 04, 2021.

VanRaden, P. M., 2008 Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414–4423.

Viteri, F. E., 1998 A new concept in the control of iron deficiency: community-based preventive supplementation of at-risk groups by the weekly intake of iron supplements. Biomed. Environ. Sci. 11: 46–60.

Von Wiren, N., S. Mori, H. Marschner, and V. Romheld, 1994 Iron inefficiency in maize mutant ys1 (*Zea mays* L. cv Yellow-Stripe) is caused by a defect in uptake of iron phytosiderophores. Plant Physiol. 106: 71–77.

Wang, Z., Z. Wang, L. Shi, L. Wang, and F. Xu, 2010 Proteomic alterations of *Brassica napus* root in response to boron deficiency. Plant Mol. Biol. 74: 265–278.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Watanabe, T., M. R. Broadley, S. Jansen, P. J. White, J. Takada *et al.*, 2007 Evolutionary control of leaf element composition in plants. New Phytol. 174: 516–523.

Waters, B. M., and R. P. Sankaran, 2011 Moving micronutrients from the soil to the seeds: genes and physiological processes from a biofortification perspective. Plant Sci. 180: 562–574.

Whitt, L., F. K. Ricachenevsky, G. Z. Ziegler, S. Clemens, E. Walker *et al.*, 2020 A curated list of genes that affect the plant ionome. Plant Direct 4: e00272.

Wolfinger, R., W. T. Federer, and O. Cordero-Brana, 1997 Recovering information in augmented designs, using SAS PROC GLM and PROC mixed. Agron. J. 89: 856–859.

Wu, D., R. Tanaka, X. Li, G. P. Ramstein, S. Cu *et al.*, 2021 High-resolution genome-wide association study pinpoints metal transporter and chelator genes involved in the genetic control of element levels in maize grain. G3 in review.

Yang, M., K. Lu, F.-J. Zhao, W. Xie, P. Ramakrishna *et al.*, 2018 Genome-wide association studies reveal the genetic basis of ionomic variation in rice. Plant Cell 30: 2720–2740.

Yıldırım, K., and S. Uylaş, 2016 Genome-wide transcriptome profiling of black poplar (*Populus nigra* L.) under boron toxicity revealed candidate genes responsible in boron uptake, transport and detoxification. Plant Physiol. Biochem. 109: 146–155.

Yu, J., G. Pressoir, W. H. Briggs, I. Vroh Bi, M. Yamasaki *et al.*, 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

Zang, J., Y. Huo, J. Liu, H. Zhang, J. Liu *et al.*, 2020 Maize YSL2 is required for iron distribution and development in kernels. J. Exp. Bot. 71: 5896–5910.

Zhang, Z., E. Ersoz, C.-Q. Lai, R. J. Todhunter, H. K. Tiwari *et al.*, 2010 Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Zhou, X., S. Li, Q. Zhao, X. Liu, S. Zhang *et al.*, 2013 Genome-wide identification, classification and expression profiling of *nicotianamine synthase (NAS)* gene family in maize. BMC Genomics 14: 238.

Zhou, X., and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods 11: 407–409.

Ziegler, G., P. J. Kear, D. Wu, C. Ziyomo, A. E. Lipka *et al.*, 2017 Elemental accumulation in kernels of the maize nested association mapping panel reveals signals of gene by environment interactions. BioRxiv 164962.

Ziegler, G., A. Terauchi, A. Becker, P. Armstrong, K. Hudson *et al.*, 2013 Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. Plant Genome 6: 2012–2007.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Table Captions

Table 1. Means and ranges for untransformed best linear unbiased predictors (BLUPs) of 15 fresh kernel elemental phenotypes evaluated in the sweet corn association panel and estimated heritability ($\hat{h}_l^2$) on a line-mean basis across two years and two locations.

Table 2. Estimated effects of endosperm mutation type from untransformed best linear unbiased predictors of 15 fresh kernel elemental phenotypes across two years and two locations.

Table 3. Predictive abilities of whole-genome prediction models for 15 fresh kernel elemental phenotypes in the sweet corn association panel.

Table S1. Lambda values used in Box-Cox transformation of 15 fresh kernel elemental phenotypes in sweet corn.

Table S2. Sources of variation for 15 elemental phenotypes in fresh sweet corn kernels from across locations, only NY, and only WI. The star (*) indicates random effect terms significant according to a likelihood ratio test ($\alpha = 0.05$).

Table S3. Means and ranges for best linear unbiased predictors (BLUPs) of 15 fresh kernel elemental phenotypes evaluated in the sweet corn association panel and estimated heritability on a line-mean basis across two years in each of two locations.

Table S4. Significant results from a genome-wide association study of 15 fresh kernel elemental phenotypes in sweet corn and their intersection with joint-linkage QTL support intervals (NAM QTL number; Table S7) of elemental phenotypes analyzed in the maize NAM panel (Ziegler *et al.* 2017).

Table S5. List of top BLASTP hits in rice and Arabidopsis for the eight maize candidate genes identified via a genome-wide association study of 15 fresh kernel elemental phenotypes in sweet corn.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Table S6. Multi-locus mixed-model results from an analysis of fresh kernel elemental phenotypes for chromosomes 2, 7, 9, and 10.

Table S7. Joint-linkage QTL support intervals (SI) of elemental phenotypes analyzed in the maize NAM panel (Ziegler *et al.* 2017) uplifted from B73 RefGen_v2 to v4.

Table S8. Genome-wide association study results of elemental phenotypes analyzed in the maize NAM panel (Ziegler *et al.* 2017) uplifted from RefGen_v2 to v4. Only NAM marker variants with resample model inclusion probability (RMIP) ≥ 5 are shown and those that reside within joint-linkage QTL support intervals (Table S7) are demarcated in the "NAM QTL number" column. The relationship of NAM marker variants to the four candidate genes identified in genome-wide association study in sweet corn that are coincident with joint-linkage QTL support intervals are presented.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

Figure Legends

Figure 1. Genome-wide association study for cadmium concentration in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNP) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the r^2 values of each SNP relative to the peak SNP (indicated in red) at 162,398,589 bp (B73 RefGen_v4) on chromosome 2. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical dashed lines indicate the genomic positions of the *heavy metal ATPase 4* (*hma4*; Zm00001d005189; 163,016,710-163,020,248 bp) and *heavy metal ATPase 3* (*hma3*; Zm00001d005190; 163,038,225-163,041,426 bp) genes. These two genes are separated by a physical distance of ~18 kb, thus their positions are not distinguishable at the plotted scale. (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The SNP from the optimal multi-locus mixed-model (S2_157751802) was included as a covariate in the mixed linear model to control for the large-effect locus. None of the tested SNPs were significant at a 5% FDR.

Figure 2. Genome-wide association study for zinc concentration in fresh kernels of sweet corn. (A) Scatter plot of association results from a mixed model analysis and linkage disequilibrium (LD) estimates (r^2). The vertical lines are $-\log_{10}$ *P*-values of single nucleotide polymorphisms (SNP) and blue color represents SNPs that are statistically significant at a 5% false discovery rate (FDR). Triangles are the r^2 values of each SNP relative to the peak SNP (indicated in red) at 180,076,727 bp (B73 RefGen_v4) on chromosome 7. The red horizontal dashed line indicates the $-\log_{10}$ *P*-value of the least statistically significant SNP at a 5% FDR. The black vertical

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

dashed line indicates the genomic position of the *nicotianamine synthase 5* (*nas5*; Zm00001d022557;179,964,493-179,965,584 bp) gene. (B) Scatter plot of association results from a conditional mixed linear model analysis and LD estimates (r^2). The SNP from the optimal multi-locus mixed-model (S7_174515604) was included as a covariate in the mixed linear model to control for the large-effect locus. None of the tested SNPs were significant at a 5% FDR.

Figure S1. Correlation matrix for untransformed, across-location BLUPs of 15 elemental phenotypes from fresh kernels in the sweet corn association panel. Pearson's correlation coefficients (r) are presented in the upper triangle, while the corresponding P-values for the significance of associations ($\alpha = 0.05$) are displayed below the diagonal.

Figure S2. Sources of variation for 15 elemental phenotypes in fresh sweet corn kernels from across locations (A), only NY (B), and only WI (C). The phenotypic variance of each trait was statistically separated into the following components: environment (Env), set within environment [Set(Env)], block within set within environment [Block(Set×Env)], genotype (Geno), genotype-by-environment interaction (Geno×Env), inductively coupled plasma mass spectrometry run (ICP), kernel sample (Sample), row within environment [Row(Env)], column within environment [Col(Env)], and residual error variance (Residual). Variance component estimates were calculated for all random effects from the full model (Equation 1).

Figure S3. Distribution of within-location BLUP values for 15 fresh kernel elemental phenotypes in the sweet corn association panel evaluated in NY and WI. Mean values of fresh kernel elemental concentration with significant differences ($P < 0.05$) between the two locations were designated with an asterisk ('*') according to a paired t-test.

Figure S4. Correlation matrix for untransformed, within-location BLUPs of 15 elemental phenotypes from fresh kernels in the sweet corn association panel. Pearson's correlation

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.19.432009; this version posted February 20, 2021. 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-ND 4.0 International license.

coefficients (*r*) between phenotypes within NY and WI are presented in the upper and lower triangles, respectively. The diagonal represents the correlation of each trait between locations. Bolded values correspond to significant correlations (*P*-value ≤ 0.05).

Figure S5. Genome-wide association study conducted across both locations for 15 fresh kernel elemental phenotypes in sweet corn. Each point represents a SNP with its -$\log_{10}$ *P*-value (y-axis) from a mixed linear model analysis plotted as a function of physical position (B73 RefGen_v4) across the 10 chromosomes of maize (x-axis). The red horizontal dashed line indicates the -$\log_{10}$ *P*-value of the least statistically significant SNP at 5% false discovery rate.

Figure S6. Genome-wide association study conducted within the New York location for 15 fresh kernel elemental phenotypes in sweet corn. Each point represents a SNP with its -$\log_{10}$ *P*-value (y-axis) from a mixed linear model analysis plotted as a function of physical position (B73 RefGen_v4) across the 10 chromosomes of maize (x-axis). The red horizontal dashed line indicates the -$\log_{10}$ *P*-value of the least statistically significant SNP at 5% false discovery rate.

Figure S7. Genome-wide association study conducted within the Wisconsin location for 15 fresh kernel elemental phenotypes in sweet corn. Each point represents a SNP with its -$\log_{10}$ *P*-value (y-axis) from a mixed linear model analysis plotted as a function of physical position (B73 RefGen_v4) across the 10 chromosomes of maize (x-axis). The red horizontal dashed line indicates the -$\log_{10}$ *P*-value of the least statistically significant SNP at 5% false discovery rate.