



Identification and validation of genomic regions influencing kernel zinc and iron in maize

Vemuri Hindu^{1,3} · Natalia Palacios-Rojas² · Raman Babu^{1,4} · Willy B. Suwarno^{2,5} · Zerka Rashid¹ · Rayalcheruvu Usha³ · Gajanan R Saykhedkar^{1,6} · Sudha K. Nair¹

Received: 17 October 2017 / Accepted: 16 March 2018 / Published online: 24 March 2018
© The Author(s) 2018

Abstract

Key message Genome-wide association study (GWAS) on 923 maize lines and validation in bi-parental populations identified significant genomic regions for kernel-Zinc and-Iron in maize.

Abstract Bio-fortification of maize with elevated Zinc (Zn) and Iron (Fe) holds considerable promise for alleviating under-nutrition among the world's poor. Bio-fortification through molecular breeding could be an economical strategy for developing nutritious maize, and hence in this study, we adopted GWAS to identify markers associated with high kernel-Zn and Fe in maize and subsequently validated marker-trait associations in independent bi-parental populations. For GWAS, we evaluated a diverse maize association mapping panel of 923 inbred lines across three environments and detected trait associations using high-density Single nucleotide polymorphism (SNPs) obtained through genotyping-by-sequencing. Phenotyping trials of the GWAS panel showed high heritability and moderate correlation between kernel-Zn and Fe concentrations. GWAS revealed a total of 46 SNPs (Zn-20 and Fe-26) significantly associated ($P \leq 5.03 \times 10^{-05}$) with kernel-Zn and Fe concentrations with some of these associated SNPs located within previously reported QTL intervals for these traits. Three double-haploid (DH) populations were developed using lines identified from the panel that were contrasting for these micronutrients. The DH populations were phenotyped at two environments and were used for validating significant SNPs ($P \leq 1 \times 10^{-03}$) based on single marker QTL analysis. Based on this analysis, 11 (Zn) and 11 (Fe) SNPs were found to have significant effect on the trait variance ($P \leq 0.01$, $R^2 \geq 0.05$) in at least one bi-parental population. These findings are being pursued in the kernel-Zn and Fe breeding program, and could hold great value in functional analysis and possible cloning of high-value genes for these traits in maize.

Introduction

Advances in agricultural research and technology have resulted in the increase of food grain production to meet the needs of an increasing human population. Though the increase in food grain production helped to meet the calorie requirement, low levels of micronutrients, zinc (Zn), iron (Fe), and pro-vitamin A among others, are a global cause for malnutrition related health impairments which could lead to

socio-economic losses, reduced work performance and productivity (Diepenbrock and Gore 2015; Tiwari et al. 2016). Micronutrient malnutrition, known as “hidden hunger”, is more prevalent among pregnant women and infants dwelling in the developing world, where people mostly rely on cereal-based diets (Diepenbrock and Gore 2015). In a specific mention about Zimbabwe, Banziger and Long (2000) reported that there was approximately 30% of pregnant and lactating women who are Fe-deficient. Although micronutrients are required in a relatively small quantity for humans, they play a vital role to stimulate cellular growth, signaling, function and metabolism (Guerrant et al. 2000; Kapil and Bhavna 2002). Micronutrients are not synthesized by humans and must be acquired through the diet. Several strategies have been used to enrich human diet with micronutrients, including supplementation, dietary diversity and commercial food fortification. However, implementing such interventions on a large scale in resource-poor communities incur substantial

Communicated by Jianbing Yan.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00122-018-3089-3>) contains supplementary material, which is available to authorized users.

✉ Sudha K. Nair
sudha.nair@cgiar.org

Extended author information available on the last page of the article

recurring expenditure and are frequently found to be economically unsustainable in the long term (Bouis and Welch 2010). Therefore, ‘bio-fortification’ which refers to micronutrient enrichment of major staple food crops like maize, rice, and wheat through classical and modern plant breeding strategies assume considerable significance (Graham and Welch 1996; Bouis 2000; Welch and Graham 2002; Babu et al. 2013).

Maize alone contributes over 20% of total calories in human diets in 21 countries and over 30% in 12 countries that are home to more than 310 million people (Smale et al. 2013). Bio-fortification of maize is, therefore, expected to improve the health and welfare of many disadvantaged populations across the globe. Significant progress has been made in developing, testing, and deploying bio-fortified maize, especially quality protein maize (QPM) (Atlin et al. 2010) and pro-vitamin A-enriched maize (Pixley 2013) worldwide. Studies have demonstrated that high-Zn maize will be impactful in rural areas with limited access to dietary supplements and fortified foods (Chomba et al. 2015). Recently, a Zn-biofortified maize variety BIO-MZN01 with 36% increased Zn on average than other maize varieties was released in Colombia in 2018. This variety was developed by International Maize and Wheat Improvement Center (CIMMYT) with the support of HarvestPlus in collaboration with the Agriculture for Nutrition Health (A4NH) and International Center for Tropical Agriculture (CIAT) (<http://www.HarvestPlus.org>).

Based on estimated average requirement (EAR) of 1860 µg/day of Zn and 1460 µg/day of Fe, the target level of Zn and Fe was established as 33 and 52 µg/g, respectively, in maize kernels (Bouis and Welch 2010). The baseline content for Zn in maize is about 20 µg/g, therefore, an increase of 13 µg/g is feasible by conventional breeding, especially because a wide range of Zn concentration is available in the germplasm. However, for Fe, such natural diversity has not been found and an increase of about 30 µg/g might be more attainable using alternate methods such as gene editing or transgenics (Ortiz-Monasterio et al. 2007).

Understanding the extent of genetic variability for kernel micronutrients in elite maize germplasm along with the genetics of accumulation mechanisms will be critical for the development of nutrient-enriched varieties. Kernel micronutrient concentration depends upon a number of factors such as micronutrient availability, environmental conditions, mobilization of nutrients from soil, uptake by roots, translocation, redistribution within the plant, and deposition in the seeds. Each of these processes is likely governed by many genes (Bashir et al. 2012; Kobayashi and Nishizawa 2012). Several genes related to metal transport, phytosiderophore biosynthesis, mineral ion sequestration and grain partitioning have been identified in Arabidopsis, rice, wheat, barley, maize, tomato and soybean (Eide et al. 1996; Zhao and

Eide 1996; Grotz et al. 1998; Eckardt 2000; Vert et al. 2001; Waters 2002). Sharma and Chauhan (2008) predicted a total of 48 candidate genes to be involved in the Fe and Zn transport in maize based on putative candidate genes, viz., 13 genes from ZIP (zinc-regulated transporter/iron-regulated transporter proteins) family, 16 from NRAMP (natural resistance associated macrophage protein) family, 17 from YS (yellow stripe) family and one each from CE (cation efflux) family and ferritin family. This makes the accumulation of minerals in seeds a complex polygenic phenomenon.

A large number of maize germplasm accessions with high levels of kernel-Zn (3.81–95.62 mg/kg) and Fe (9.6–159.43 mg/kg) have been identified in temperate (Ahmadi et al. 1993; Brkic et al. 2004) and tropical collections (Banziger et al. 2000; Menkir 2008; Chakraborti et al. 2011; Prasanna et al. 2011), in mid-altitude and low-land agro-ecologies including landraces, inbreds, hybrids and open pollinated varieties. Genome analysis tools provide access to thousands of genomic polymorphisms, thus considerably broadening our capacity to monitor and effectively utilize genetic diversity (Glaszmann et al. 2010). Genome-wide association studies (GWAS) based on linkage disequilibrium (LD) is a robust approach for mapping biologically valuable traits in germplasm and has been successfully applied in a range of plant species (Huang and Han 2014; Yang et al. 2014). LD is the non-random association of alleles at different loci, measured as r^2 and D' (Flint-Garcia et al. 2003). Genotyping-by-sequencing (GBS) methodology (Elshire et al. 2011) offers a less expensive method for genotyping large number of samples, and provides around a million SNPs, and hence is extremely relevant to LD-based mapping in a crop like maize which has reported faster LD decay. Validated marker-trait associations from GWAS will be of great value in developing bio-fortified maize with high kernel-Zn and Fe. Objectives of the current study were (1) to assess the genetic variation for kernel-Zn and Fe concentrations in a wide array of maize germplasm; (2) to identify the genomic regions that influence kernel-Zn and Fe concentrations through GWAS and develop high throughput and easy to use SNP assays; and (3) to validate GWAS-identified genomic regions in bi-parental populations.

Materials and methods

Plant materials and growing environments

A set of 923 inbred lines representing CIMMYT and partners’ germplasm was used as an association mapping panel. The panel included 432 tropical, 402 subtropical and 89 temperate germplasm. In addition to elite breeding lines, a number of lines from improved pools and

populations formed to serve as sources for drought and Low N tolerance were included, as well as lines that were developed for QPM.

This panel was grown in three different environments at CIMMYT research stations in Mexico: at Agua Fria in 2012 (AF12A) and 2013 (AF13A) and at Celaya in 2012 (CE12B). Agua Fria is located at 20°32'N, 97°28'W, 110 m above sea level (masl), and has average annual temperature of 22 °C with average annual precipitation of 1200 mm. Celaya is located at 20°26'N, 103°19'W; 1750 masl; average annual temperature 19 °C; average annual precipitation 700 mm. The trials at AF12A and CE12B were laid out in a randomized complete block design (RCBD) with two replications, whereas at AF13A, in alpha lattice design with two replications. The rows were of 2.5 m length and 75 cm apart and each entry was grown in a single row plot. Two to six plants from each plot were self-pollinated, hand-harvested and hand-shelled to avoid any metal contamination. Kernels were bulked for subsequent micronutrient analyses. Soil samples were also taken in at least five distal points in the field trials and at 0–30 cm deep. Samples were dried and sent for analysis at Fertilib commercial laboratory in Mexico.

From the association mapping panel, three lines with high Zn and three lines with moderate or low Zn lines were selected based on the micronutrient analysis to form the bi-parental populations. Three double-haploid (DH) populations were derived from the crosses between high Zn lines with low or moderate Zn lines (Table 2). The three DH populations (DHP1, DHP2 and DHP3) had population sizes of 96, 112 and 143, respectively. These populations were planted in single replication trials in two environments at Celaya in 2014 and in Tlatizapan, Morelos, Mexico in 2015. Planting conditions and micronutrient analysis methods were the same as described above.

Micronutrient analysis

Random samples of 6 g were used for analysis. Only six kernels were ground into fine powder (<0.5 µm), using a Retsch™ miller (model MM400) and 35 mL grinding milling jar of zirconium. Milling time was 90 s at 30 Hz. Flour was collected in 15 mL plastic tubes and analyzed by X-ray fluorescence using X-ray fluorometer (XRF) Oxford instruments™, model X-Supreme 8000®. Five grams of flour was placed in the polypropylene capsules and closed with a Poly-4® Oxford Instruments™, and readings were recorded. About 10% of the samples were also analyzed by inductive coupled plasma (ICP) as described by Galicia et al. (2012) to confirm accuracy of values obtained by XRF. In ICP

analysis, aluminum and titanium were also monitored as indicators of contamination.

Genotyping

DNA was extracted from leaf samples of 3–4-week-old seedlings using the standard CIMMYT laboratory protocol (Cimmyt 2005). The association mapping panel and three DH populations under study were genotyped for single nucleotide polymorphism (SNP) using genotyping-by-sequencing (GBS) method at the Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA. Physical coordinates of all SNPs were derived from the maize reference genome version B73 AGPV2. The genotypic data consisted of 955,690 SNPs across all the chromosomes, in the imputed GBS SNP dataset of approximately 22,000 maize samples publicly available through Panzea (<http://www.panzea.org>). From this, a smaller dataset of 347,765 SNPs which met the filtering criteria of call rate (CR) ≥ 0.7 and minor allele frequency (MAF) ≥ 0.03 was used for GWAS. For principal component and kinship analyses, 69,830 SNPs with filtration criteria of CR ≥ 0.9 and MAF ≥ 0.1 were used.

Phenotypic data analysis

Variance components, σ^2G , σ^2GE and σ^2e , for the multi-environment phenotypic data were estimated from analysis of variance (ANOVA) using multi environment trial analysis with R (METAR) (Alvarado et al. 2015). Broad-sense heritability (H^2) of the trials was estimated as:

$$H^2 = \frac{\sigma^2G}{\sigma^2G + \sigma^2GE/l + \sigma^2e/lr},$$

where σ^2G is the genotypic variance, σ^2GE is the genotype × environment variance, σ^2e is the error variance, l is the number of environments, and r is the number of replications. Correlation coefficients between environments and traits, summary statistics (mean, SE, range, LSD, CV) were also generated using standard procedures implemented in METAR. Best linear unbiased estimators (BLUEs) used for GWAS was estimated as

$$Y_{ijkl} = \mu + \text{Loc}_l + \text{Rep}_j(\text{Loc}_l) + \text{Block}_k(\text{Rep}_j\text{Loc}_l) + \text{Gen}_i + \text{Loc}_l \times \text{Gen}_i + \varepsilon_{ijkl},$$

where Y_{ijkl} is the response value of observed trait, μ the overall mean, Gen_i is the treatment fixed effect ($i = 1, 2, \dots, n$), Rep_j is the replicate effect ($j = 1, 2, \dots, n$), Block_k is the block effect, Loc_l is the location effect and ε_{ijkl} is the error term.

Population structure, kinship and genome-wide linkage disequilibrium

The principal component analysis (PCA) method as described by Price et al. (2006), implemented in SNP and Variation Suite (SVS) V_8.6.0 (SVS, Golden Helix, Inc., Bozeman, MT, USA, <http://www.goldenhelix.com>) was used for the analysis. A three-dimensional plot of the first three principal components was drawn to visualize the possible population stratification among the samples. A kinship matrix was also computed from identity-by-state (IBS) distance matrix as executed in SVS V_8.6.0:

IBS distance =

$$\frac{\text{No. of markers IBS2} + (0.5 \times \text{No. of markers in IBS1})}{\text{No. of non-missing markers}},$$

where IBS1 and IBS2 are the states in which the two inbred lines share one or two alleles, respectively, at a marker (Bishop and Williamson 1990).

Linkage disequilibrium (LD) was quantified as adjacent-pairwise r^2 values (the squared allele frequency correlations, among alleles at two adjacent SNP markers) (Hill and Robertson 1968) and was estimated for 34,420 SNPs using SVS V_8.6.0. To investigate the extent of linkage disequilibrium (LD) decay across the genome, r^2 values were plotted against the physical distance within the SNPs (Remington et al. 2001). The ‘nlin’ function in the statistical programming language R (R Core Team 2017) was used to obtain LD decay plot as non-linear model.

GWAS for kernel-Zn and Fe

For each trait, three different association analyses were carried out: uncorrected (U), corrected for population structure (Q), and corrected for population structure and kinship ($Q+K$) using SVS V_8.6.0. In the uncorrected analysis, associations were tested in an additive model without correcting for any of the confounding variables. In the Q model (GLM—general linear model, or fixed-effect linear model), the associations were corrected using population structure through principal component analysis, in which ten principal components (PC) were included. In the $Q+K$ model (MLM—mixed linear model), associations were corrected using both PCs and kinship matrix. All the three models (U , Q and $Q+K$) involved testing one variant at a time. Manhattan plots were plotted using the $-\log_{10} P$ values of all SNPs used in analysis. The appropriateness of the different models was evaluated through $Q-Q$ plots that were obtained by plotting ‘expected $-\log_{10} P$ values’ on the x -axis and ‘observed $-\log_{10} P$ values’ on the y -axis. Multiple testing correction was performed to determine the significance threshold, where instead of 345,767 independent tests, the total

number of tests were estimated based on the average extent of LD at $r^2=0.1$ (Cui et al. 2016). Based on this, significant associations were declared when the P values in independent tests are less than 5.03×10^{-05} or $-\log_{10} (P \text{ values})$ are greater than 4.3. The variance component based on the kinship analysis employing 347,765 SNPs was computed using the efficient mixed model analysis (EMMA) (Kang et al. 2008) as implemented in SVS V_8.6.0. Narrow sense heritability (pseudo-heritability) was estimated as

$$ph = \frac{\widehat{\sigma}_g^2}{\text{Var}(y)} = \frac{1}{(1 + \widehat{\delta})},$$

where $\widehat{\sigma}_g^2$ is the estimated genetic variance, $\text{Var}(y)$ is variance of observed phenotypes, $\widehat{\sigma}_e^2$ is the estimated residual variance, $\widehat{\delta}$ is $\widehat{\sigma}_e^2 / \widehat{\sigma}_g^2$.

Genes co-localized with associated SNPs were identified from the maize GDB genome browser (<http://www.maizegdb.org>) annotations were obtained from http://ensembl.gramene.org/Zea_mays.

Validation in bi-parental populations

SNPs found to be significantly associated with kernel-Zn and Fe were selected for single marker QTL analysis in three DH bi-parental populations. In addition, SNPs within the bottom 0.1 percentile of the distribution in the GWAS (Sehgal et al. 2017) were also tested independently for possible linkage to kernel-Zn and Fe concentrations in the DH populations phenotyped at two environments. Single-marker QTL analysis was carried out on single location data, from both the locations obtained from each DH population separately using Gen Stat 14.0.

Results

Phenotypic data analysis

Soil characteristic and composition were slightly different between samples from Agua Fria (AF) and Celaya (CE). Soil from CE was of clay type with a pH of 8.03, Zn content of 2.32 $\mu\text{g/g}$ dry weight (DW), Fe 7.84 $\mu\text{g/g}$ DW and N 6.15 $\mu\text{g/g}$ DW. AF soil was of clay loam type with pH of 8.28, Zn content of 1.24 $\mu\text{g/g}$ DW, Fe 17.2 $\mu\text{g/g}$ DW and N 7.57 $\mu\text{g/g}$ DW. ANOVA and other descriptive statistics for both the traits showed significant variability for kernel-Zn and-Fe concentrations among the genotypes of the association mapping panel (Table 1). The average kernel-Zn in the panel was 27.04 $\mu\text{g/g}$ DW, with a range of 17.11–43.69 $\mu\text{g/g}$ DW. The average Fe concentration was 14.65 $\mu\text{g/g}$ DW with

Table 1 Estimates of mean, variance components (across 3 locations) and broad-sense heritability for kernel-Zn and Fe concentrations in GWAS panel

Trait	Environments	Zn ($\mu\text{g/g}$)	Fe ($\mu\text{g/g}$)
Mean	AF12A	26.28 \pm 2.61	16.29 \pm 1.81
	CE12B	25.26 \pm 2.36	13.18 \pm 1.68
	AF13A	29.6 \pm 3.22	14.48 \pm 2.04
	Across	27.04 \pm 2.76	14.65 \pm 1.85
Range	AF12A	15.20–42.50	8.25–28.60
	CE12B	14.76–50.15	5.65–25.19
	AF13A	17.10–39.80	7.12–27.60
	Across	17.11–43.69	8.19–25.65
LSD	AF12A	5.11	3.54
	CE12B	4.62	3.29
	AF13A	6.30	3.99
	Across	5.41	3.63
CV	AF12A	9.91	11.10
	CE12B	9.34	12.73
	AF13A	10.87	14.07
	Across	10.22	12.64
H^2	AF12A	0.78	0.83
	CE12B	0.83	0.78
	AF13A	0.76	0.77
	Across	0.86	0.83
σ^2G		1.28***	4.65***
σ^2GE		2.12***	10.84***

AF12A Agua Fria 2012, CE12B Celaya 2012, AF13A Agua Fria 2013, H^2 broad-sense heritability, LSD least significant difference, CV coefficient of variance, G genotype, E environment

***Indicates significance at 0.001

a range of 8.19–25.65 $\mu\text{g/g}$ DW. None of the lines in the association mapping panel met the target kernel-Fe concentration of 52 $\mu\text{g/g}$. Correlations between environments were highly significant for both kernel-Zn and Fe (Table S1). Highly significant, but moderate positive correlation was found between the two traits across the environments ($r=0.49$, P value ≤ 0.001). In the GWAS panel, there were 57 lines (Table S2) which had Zn concentration of 33 $\mu\text{g/g}$ DW or more, the current breeding target for HarvestPlus. Among those 57, there are several elite CIMMYT maize lines (CMLs), 4 QPM lines, several lines susceptible or tolerant to drought, heat or a combination of drought and heat as well as lines susceptible or tolerant to low nitrogen (low N).

Principal component analysis and genome-wide linkage disequilibrium

Principal component analysis using genome-wide markers revealed only a moderate population structure with the first three principal components (Fig. S1). The temperate lines and the drought tolerant donor germplasm including both

La Posta Sequia and drought tolerant population (DTP) groups clearly separated in different axes from the rest of the CIMMYT tropical and sub-tropical lines. The first three principal components explained 41.86% of the total variance. The genome-wide LD decay was plotted as LD (r^2) between adjacent pairs of markers versus distance in kb between adjacent pairs, and showed that the average LD decay was 17.5 kb at $r^2=0.1$ and 5.99 kb at $r^2=0.2$ (Fig. S2). Chromosome-wise LD analyses showed the slowest LD decay on chromosome 8 (26.54 kb, $r^2=0.1$), followed by chromosome 1 (21.88 kb, $r^2=0.1$).

GWAS for kernel-Zn and Fe

GWAS was carried out with a subset of SNPs with call rate (CR) more than 0.7 and minor allele frequency (MAF) more than 0.03. 347,765 SNPs which formed this subset were tested against kernel-Zn and Fe concentrations from 923 inbred lines across environments. In the three models used for GWAS, SNPs were tested independently against the phenotypes for Zn and Fe. The uncorrected method (U) and the method corrected only for population structure (Q) showed significant genomic inflation as observed in the $Q-Q$ plots (Fig. 1). The $Q+K$ MLM, where individual SNPs were tested independently correcting for both population structure and kinship, showed the least genomic inflation for both Zn and Fe, and hence significant associations were finalized based on this analysis (Fig. 1). The narrow sense heritability for kernel-Zn was estimated as 0.72 based on the IBS kinship matrix employing all SNPs used in GWAS, with a standard error of 0.16. A total of 20 SNPs were found to have a significant association with kernel-Zn with the P value range from 4.93×10^{-06} to 5.03×10^{-05} (Fig. 2; Table S3) and S4_843764 and S4_843777 on chromosome 4 were found to be the most significantly associated SNPs to kernel-Zn in the panel. Among the 20 SNPs identified for kernel-Zn, 14 were located within predicted gene models, of which 5 were within models with functional domains generally related to metal ion binding or transport or specifically to Zn ion binding (Table S3). Four SNPs identified from the GWAS were found to be located within previously reported QTL for kernel-Zn in maize (Table 4).

For kernel-Fe, a narrow sense heritability of 0.70 was estimated with a standard error of 0.27. 26 SNPs were found to be significantly associated with kernel-Fe with P values ranging between 2.43×10^{-06} and 5.03×10^{-05} (Fig. 2; Table S3), with the maximum number of SNPs found on chromosome 1 (eight SNPs). S1_64238426 on chromosome 1 and S9_136390177 on chromosome 9 were found to be the most significantly associated SNPs with kernel-Fe in this panel. The proportion of variance explained by individual SNPs ranged from 1.8 to 2.41%. Among the 26 SNPs, 20 SNPs were located within predicted gene models. Seven of

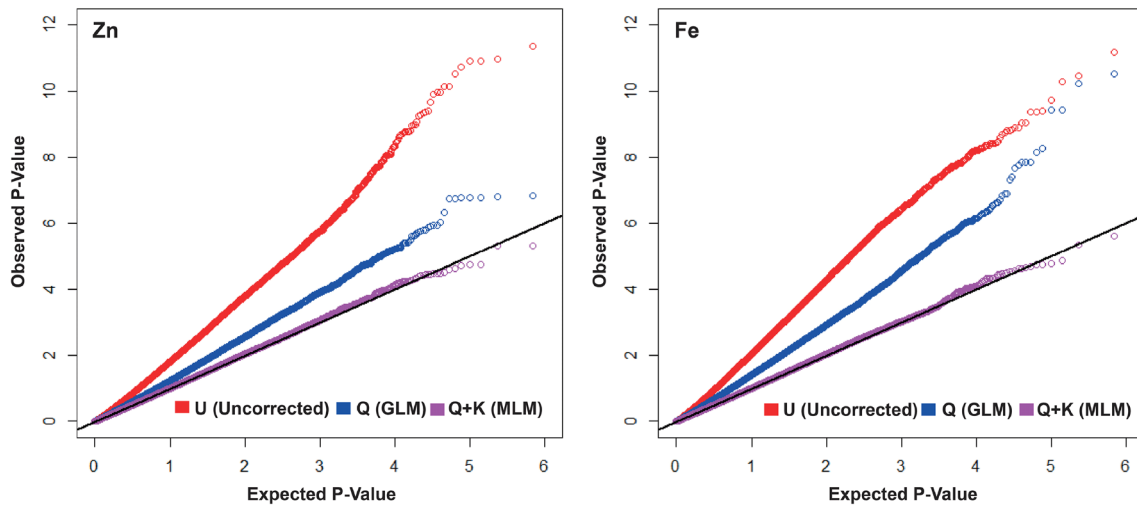
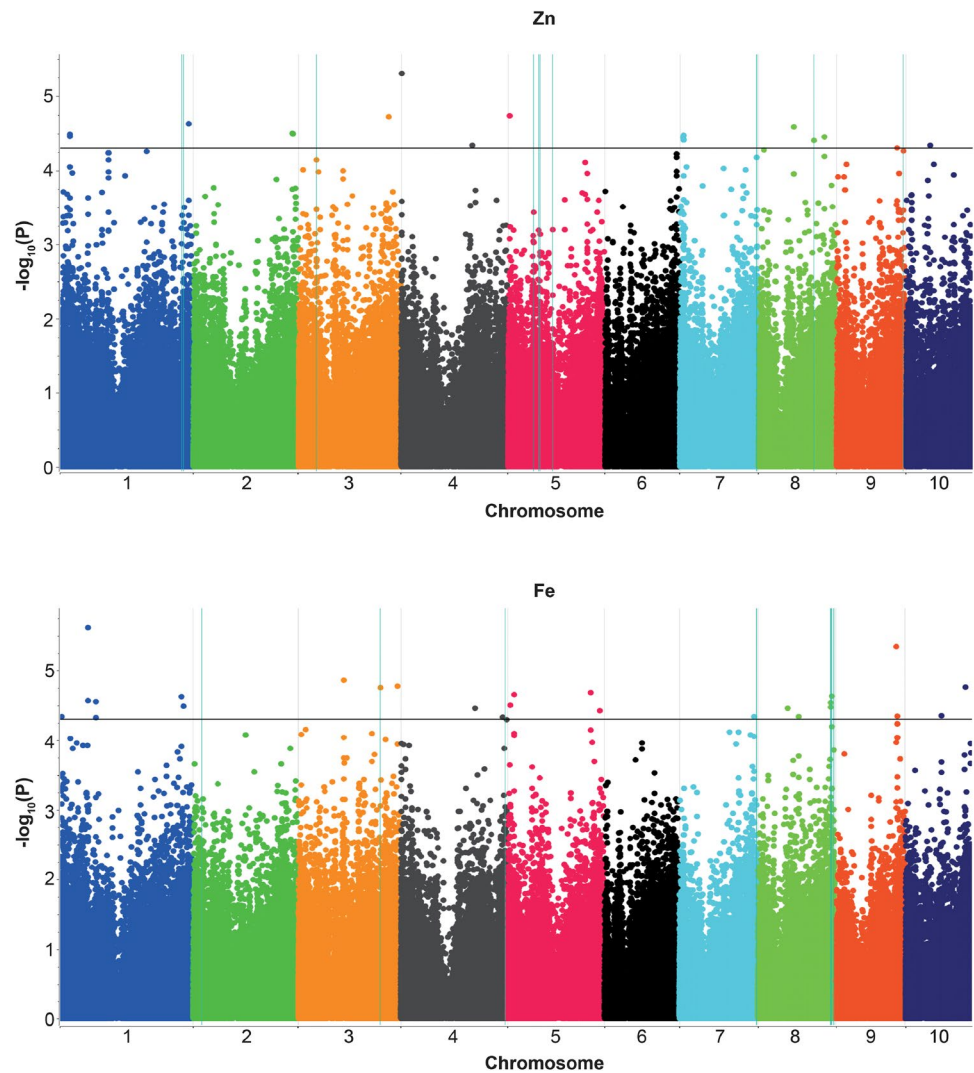


Fig. 1 Quantile–quantile (Q – Q) plots showing inflation of estimated $-\log_{10}(P)$ values versus observed for traits Zn and Fe using uncorrected association model (U), Q (GLM) and $Q+K$ (MLM). Q ten

principal components (fixed), K kinship matrix (random), GLM general linear model, MLM mixed linear model

Fig. 2 Manhattan plot from the $Q+K$ (MLM) model for Zn and Fe, plotted with the individual SNPs of all chromosomes on the X -axis and $-\log_{10}P$ value of each SNP in the Y -axis. The different colors indicate the 10 chromosomes of maize. The black horizontal line shows the multiple testing threshold $-\log_{10}P$ value of 4.3 for this panel. The blue vertical lines show the associated SNPs validated in bi-parental populations



the GWAS SNPs were located within QTLs reported for kernel-Fe in maize (Table 4).

A set of 57 lines with the highest kernel-Zn content of > 33 µg/g were compared against the same number of lines which had the least kernel-Zn content in the panel. There was a clear enrichment of the favorable allele in all the Zn-associated SNPs in the 57 lines with high kernel-Zn ranging from 51.02 (S10_54119964) to 90.91% (S2_225529232). Similar analysis with high and low kernel-Fe lines showed an allele enrichment of favorable alleles in the high kernel-Fe lines ranging from 52.31% (S10_136070835) to 93.33% (S1_64238509).

Validation in bi-parental populations

Three DH populations that were phenotyped for kernel-Zn and Fe showed considerable range for the two traits in the two environments studied (Table 2). The kernel-Zn ranged from 15.6 and 48.0 µg/g DW across the two environments and three populations, similarly kernel-Fe ranged between 6.3 and 24.5 µg/g DW. DHP2 showed wider range of concentrations for both kernel-Zn and Fe. Other than the 20 SNPs for kernel-Zn and 26 SNPs for kernel-Fe identified based on GWAS *P* values lower than the panel-determined threshold, 381 SNPs were selected for kernel-Zn and 345 SNPs for kernel-Fe under a reduced threshold limit of *P* value $\leq 1.0 \times 10^{-03}$, for single marker QTL analysis. From these, the polymorphism between the respective parents of each DH population reduced the number of SNPs tested to 232 and 231 SNPs for Zn and Fe, respectively. These SNPs were tested for linkage to kernel-Zn and Fe concentration independently in the three bi-parental populations by analysis of variance due to each allele class in the DH populations. This analysis identified 11 SNPs each for kernel-Zn and Fe that had significant effect on the trait variance ($P \leq 0.01$, $R^2 \geq 0.05$), in one or more populations in one or more environments (Table 3; Fig. 3). Among the 11 SNPs validated for kernel-Zn, six genomic regions could be identified. These were represented by one SNP each on chromosomes 3, 8 and 9, two SNPs each on chromosomes 1 (4 Mb interval) and 7 (adjacent SNPs) and four SNPs on

chromosome 5. Among the 11 SNPs validated for kernel-Fe, five broad genomic regions could be identified, with one SNP each from chromosomes 2, 3, 4 and 7, and seven SNPs from chromosome 8, spread in a physical interval of 8 Mb in DHP1 and DHP2. Notable among these were adjacent markers, S7_173181688 (Chr 7: 173,181,688) and S7_173181689 (Chr 7: 173,181,689) that explained 29% (LOD: 9.58) and 28% (LOD: 9.58) of the phenotypic variance for kernel-Zn in DHP3 (Table 3; Fig. 3). Similarly, S8_167013673 (chr 8: 167,013,673) explained 34% phenotypic variance (LOD: 7.72) for kernel-Fe in DHP2, along with two SNPs, S8_164741044 (chr 8: 164,741,044) and S8_164741133 (Chr 8: 164, 74, 1133) which explained about 27% variance (LOD: 7.47) in DHP2 (Table 3; Fig. 3).

Discussion

Kernel-Zn and Fe in the mapping panel

In most parts of the maize-growing areas, soils may have different chemical and physical characteristics that can significantly reduce the availability of Zn to plant roots (Cakmak 2008). Hence, the objective of the bio-fortification breeding programs is to develop cultivars that express maximum possible genetic potential to absorb sufficient Zn from the soil and accumulate it in the grain. Zinc availability is highly dependent on pH. Both the soils at CE and AF were slightly alkaline which usually leads to very low Zn availability. Nevertheless, there were lines with Zn concentration as high as 43.7 µg/g DW identified in the mapping panel, and 57 lines had kernel-Zn concentration above the required target. The genotypic variability was also high for the traits making this an ideal population set to study these traits. Among the 57 lines that have higher concentration of Zn compared to the current breeding target, 6 are CMLs (elite CIMMYT maize lines) including CML166, CML192, CML264, CML323, CML361, CML421, which have already been used in hybrid varieties, and no information was previously known about their nutrition content. Interestingly, two of the lines including an elite CML (CML361) are also acid tolerant lines.

Table 2 Pedigree and summary statistics of DH populations employed in bi-parental mapping at two locations

DH population	Pedigree	Population size	Trait	Celaya(CE)		Tlaltizapan(TZ)	
				Mean (µg/g)	Range (µg/g)	Mean (µg/g)	Range (µg/g)
DHP1	DTPYC9-F13-2-1-1-2-B-B/CML312	96	Zn	25.61 ± 0.40	17.85–37.45	26.07 ± 0.45	18.50–36.90
			Fe	12.23 ± 0.28	7.10–24.50	15.14 ± 0.35	8.40–22.50
DHP2	CML503/CLWN201	112	Zn	25.32 ± 0.47	15.80–48.00	24.01 ± 0.38	16.15–35.35
			Fe	13.67 ± 0.32	6.30–23.45	12.96 ± 0.30	7.30–22.00
DHP3	CML 465/CML451	143	Zn	27.96 ± 0.39	17.70–43.15	24.08 ± 0.33	15.60–37.80
			Fe	13.91 ± 0.20	9.00–20.70	14.26 ± 0.18	8.20–19.00

Table 3 GWAS identified markers validated in DH populations for kernel-Zn and Fe concentration

Trait	SNP	Chr	Physical Position (bp)*	GWAS_P value	DHP_P value	Add effect (μg/g)_DHP	ANOVA_adj R ² (%)	DH population	Effect_location	Gene model	Annotation
Zn	S1_275780297	1	275,780,297	6.12E-04	7.01E-04	-1.53	10.3	DHP2	CE	GRMZM2G472171	Protein_coding
	S1_279913529	1	279,913,529	5.14E-04	4.78E-03	-1.57	10.7	DHP2		GRMZM2G400714	Zinc ion binding and ZOS8-04—C ₂ H ₂ zinc finger protein
	S3_40522792	3	40,522,792	7.28E-05	9.63E-03	-0.82	5.1	DHP3	TZ	GRMZM2G089959	PPR repeat containing protein
	S5_56793531	5	56,793,531	9.09E-04	4.72E-04	1.64	10.7	DHP2	CE and TZ	GRMZM2G116586	UTP-glucose-1-phosphate uridylyltransferase
	S5_68423957	5	68,423,957	6.45E-04	2.79E-03	1.38	7.9	DHP3		GRMZM2G074896	Glycosyltransferase protein
	S5_71718466	5	71,718,466	7.35E-04	4.49E-06	1.81	13.1	DHP1		GRMZM2G144282	Translation initiation factor eIF3 subunit
	S5_100070727	5	100,070,727	6.40E-04	2.27E-06	1.60	13.3	DHP3		GRMZM2G079653	Ice binding, Homiothermy,
	S7_173181688	7	173,181,688	5.44E-04	5.08E-09	-2.24	29.2	DHP3		GRMZM2G158162	Zinc ion binding and B3 DNA binding domain containing protein
	S7_173181689	7	173,181,689	6.75E-05	5.21E-09	-2.20	28.4	DHP2		GRMZM2G048200	GDSL-like lipase,acylhydrolase
	S9_151265550	9	151,265,550	5.51E-05	1.68E-03	1.13	6.6	DHP3		GRMZM2G178775	Putative ribonuclease P
Fe	S2_19265861	2	19,265,861	9.37E-04	1.45E-04	-0.78	11.0	DHP3	CE	GRMZM2G162333	Pectinesterase activity
	S3_186200393	3	186,200,393	1.78E-05	3.23E-03	-0.88	10.1	DHP1		GRMZM2G312201	Regulation of transcription and pentatricopeptide
	S4_236412442	4	236,412,442	9.06E-04	2.23E-04	-1.14	13.9	DHP2	TZ	GRMZM2G471479	-
	S7_174289806	7	174,289,806	8.56E-04	1.97E-03	-0.90	11.0	DHP1	CE	GRMZM2G130950	5'-AMP-activated protein kinase beta-2 subunit protein
	S8_164741044	8	164,741,044	2.89E-05	4.03E-04	1.69	27.4	DHP2		GRMZM2G472991	Serine/threonine-protein kinase
	S8_164741133	8	164,741,133	3.38E-05	2.26E-07	1.71	27.7	DHP2		GRMZM2G172032	Isoprenoid biosynthetic process
	S8_164750150	8	164,750,150	6.59E-04	1.01E-04	1.19	14.8	DHP2		GRMZM2G074462	Starch binding domain containing protein
	S8_167013673	8	167,013,673	9.99E-04	2.18E-07	1.80	34.1	DHP2		GRMZM2G162329	nsp1-like C-terminal region family protein
	S8_167086477	8	167,086,477	2.36E-05	4.01E-04	1.14	12.6	DHP2		GRMZM2G061043	nsp1-like C-terminal region family protein
	S8_167089929	8	167,089,929	4.97E-04	6.34E-04	-1.15	12.9	DHP2		GRMZM2G004690	Protein_coding
	S8_172006274	8	172,006,274	1.38E-04	6.44E-06	1.47	19.2	DHP2			

The chromosomal locations along with effect estimates in different bi-parental populations, along with predicted gene models (B73 AGPV2) within which the SNPs were located are depicted

*Physical positions were retrieved from maize B73 AGPV2. CE and TZ represent the two environments Celaya and Tlalitzapan respectively

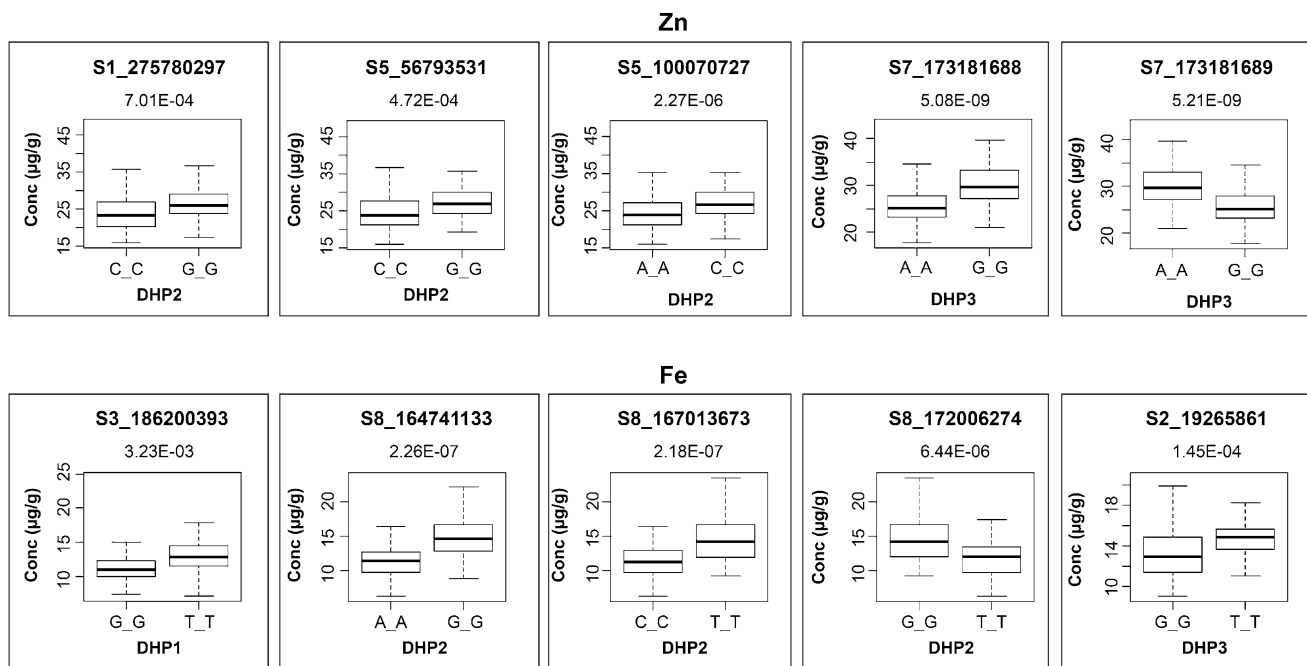


Fig. 3 Box plots showing the phenotypic values of the different allele classes of major-effect SNPs validated in DH populations for Zn and Fe (upper and lower panel, respectively). The SNP names, *P* value, alleles and the specific DH population where the effect is witnessed are mentioned near each box. The black horizontal lines in the middle

of the boxes are the median values for the Zn or Fe concentration in the respective allele classes. The vertical size of the boxes represents the inter-quantile range. The upper and lower whiskers represent the minimum and maximum values of data

Table 4 GWAS identified SNPs found within the previously reported QTLs for kernel-Zn and Fe

Trait	GWAS-identified SNPs	Chr	Physical position (bp) ^a	QTL interval	QTL bin	QTL study
Zn	S3_40522792	3	40,522,792	umc1504–umc1386a	3.04	Qin et al. (2012)
Zn	S4_161165956	4	161,165,956	bnlg1621a–dupssr16	4.06	Jin et al. (2013)
Zn	S5_100070727	5	100,070,727	umc1110–bnlg1208	5.04	
Zn	S9_151265550	9	151,265,550	umc1310–bnlg128	9.06–9.07	
Zn	S9_151265550	9	151,265,550	dupssr29–bnlg619	9.07	Qin et al. (2012)
Fe	S2_19265861	2	19,265,861	umc1542–umc1042	2.07	Qin et al. (2012)
Fe	S2_19265861	2	19,265,861	bnlg1690–umc1890	2.04–2.07	Jin et al. (2013)
Fe	S3_186200393	3	186,200,393	mmp144a–umc1266	3.04–3.06	
Fe	S4_167189737	4	167,189,737	bnlg1621a–dupssr16	4.06	
Fe	S5_5104719	5	5,104,719	rz87, RZ87	5.1	Lung’aho et al. (2011) and Baxter et al. (2013)
Fe	S8_164741044	8	164,741,044	ZM0825	8.06	Simic et al. (2012)
Fe	S8_164741133	8	164,741,133			
Fe	S8_164750150	8	164,750,150			

The physical coordinates of the GWAS-identified SNPs and chromosomal bins of markers reported in earlier QTL mapping studies are obtained from B73 AGPV2 and maize GDB

^aPhysical positions were retrieved from maize B73 AGPV2

Among the mechanisms of alleviating aluminum (Al) toxicity, chemical exudates including organic acids, phenolic compounds and phosphates prevent Al from entering to the roots and accumulating in cells (Panda et al. 2009). Some of

these mechanisms are common to mineral uptake processes, and could affect Zn uptake from soil.

Four lines identified with high kernel-Zn concentration are QPM (Table S2). An above-average concentration of

kernel-Zn was reported in the QPM germplasm as compared to non-QPM/normal maize germplasm (Chakraborti et al. 2009, 2011). In QPM, the presence of opaque-2 allele partially inhibits zein synthesis, with proportional increase in other protein fractions like glutelins, albumins, globulins, proteins known to bind Zn in the endosperm of maize (Diez-Altare and Bornemisza 1967). In addition, Zn plays an important role in tryptophan biosynthesis, which is increased in QPM. In fact, as a part of the breeding program for high Zn at CIMMYT, most of the high zinc germplasm identified so far is QPM, although not all QPM germplasm is high in Zn (Palacios-Rojas, unpublished). Out of the 923 lines used in this study, only 31 were QPM or had QPM background and 33.3% had Zn values higher than 30 $\mu\text{g/g}$ DW. In contrast, out of the 892 non-QPM used in the panel, 19.9% had values higher than 30 $\mu\text{g/g}$ DW, and about 6% of them had values higher than the breeding target (33 $\mu\text{g/g}$ DW). Taken together, these results indicate great potential to develop high Zn maize alone or in combination with better protein quality in bio-fortification programs.

Genetic control for abiotic stresses like drought, heat and a combination of these stresses are largely independent (Cairns et al. 2013) and metabolite responses have highlighted the importance of photorespiration and raffinose family oligosaccharide metabolism (Obata et al. 2015). Under drought and combined drought and heat stress, tryptophan accumulation in maize leaves has been reported in the susceptible genotypes (Obata et al. 2015; Witt et al. 2012). Among the lines with higher Zn concentration, ten genotypes were susceptible to drought or to combined drought and heat stress (Table S2), which is in accordance to the role that Zn plays during tryptophan biosynthesis. However, one genotype tolerant for drought and two tolerant for combined drought and heat stress also had high values of Zn, which underlines the involvement of Zn in other pathways and provides opportunities to develop high Zn germplasm that could be tolerant to such abiotic stresses. Interestingly, four genotypes with high Zn values have been previously reported as susceptible to low-N. It will be important to screen the kernel-Zn accumulation in germplasm exposed to abiotic conditions like drought, heat, low-N, low phosphorous and combined stresses. There is a need to assess if there has been selection against traits like Zn, or if stress tolerance and kernel-Zn can be combined—this is particularly important as drought, heat and combination of stresses are going to become increasingly prevalent under climate change in many countries where mineral deficiency in the diet is prevalent. Equally important is to understand kernel-Zn accumulation in acid-soil tolerant germplasm. This could open the possibility to develop germplasm tolerant to acid soil and capable to accumulate high Zn, which could be ideal products for HarvestPlus target countries like Colombia.

The kernel-Zn and Fe showed significant, but moderate positive correlation in the association mapping panel ($r = 0.49$, $P \leq 0.001$), which was similar to some earlier reports (Maziya-Dixon et al. 2000; Lung'aho et al. 2011; Baxter et al. 2013) in maize. A significant correlation between grain Zn and Fe concentrations was also reported in wheat (Velu et al. 2011), rice (Kabir et al. 2003), pearl millet (Velu et al. 2008; Gupta et al. 2009) and sorghum (Kumar et al. 2009). This suggests that these traits might have some common genetic mechanisms leading to their accumulation in grains. For instance, some common members of the ZIP family, which is involved in the transport of Zn and Fe as well as of other varieties of divalent cations have been reported (Lee et al. 2010). In addition, several genes responsible for metal chelation, phytosiderophore biosynthesis, uptake, transport, loading and storage of these minerals have been identified in rice, barley, wheat and maize (Gross et al. 2003; Anuradha et al. 2012; Bashir et al. 2012; Sharma and Chauhan 2008).

Population structure and linkage disequilibrium in the mapping panel

The panel with 923 inbred lines showed moderate population structure within it, based on the principal component analysis. This panel, as discussed before, had germplasm from other breeding programs apart from CIMMYT's sub-tropical and tropical germplasm. CIMMYT's elite germplasm showed less spread along the axes, and this has been observed in other studies also, where association mapping panels were constituted only with CIMMYT sub-tropical and tropical lines. In some studies, the first three PCs, explained only about 20% of the total variation (Rashid et al. 2018). Warburton et al. (2002) also observed that the CIMMYT populations, from which most of the sub-tropical and tropical lines have been derived, had a large amount of diversity within, rather than between source populations. Due to the heterogeneous nature of CIMMYT populations, they suggested that it would be difficult to find well-defined structure within CIMMYT lines. The moderate structure that was observed in the present study panel may be due to the inclusion of multiple sources of germplasm, whether from the temperate breeding pools from South Africa or the drought tolerant donor lines from CIMMYT, like LaPosta Sequia and DTP lines. The macro-structure relationship within the panel and the cryptic relatedness due to kinship could confound association mapping (Yu et al. 2009), and hence the need to be using appropriate models to control spurious associations while conducting GWAS. Linkage disequilibrium (LD) is a measure of non-random association of alleles at two or more loci; faster the LD decay, better is the mapping resolution. We found an average LD decay ($r^2 = 0.2$) across the whole genome close to 6 kb in the panel. This is in accordance with

several previous studies (Yan et al. 2009; Lu et al. 2011; Romay et al. 2013; Zhang et al. 2016) employing tropical and sub-tropical maize. This rapid LD decay in the panel is reflective of the genetic diversity of the tropical/subtropical maize germplasm used in this study. Among the ten chromosomes, chromosome 8 was found to have the slowest LD decay (26.54 kb), and this was observed in several previous studies (Suwarno et al. 2014; Rashid et al. 2018).

GWAS and validation in bi-parental populations

Unlike linkage mapping, association mapping can explore all the recombination events and mutations in a given population and with a higher resolution (Yu and Buckler 2006). Population structure and cryptic relatedness in the form of kinship can create unexpected LD between unlinked loci across the genome. Many statistical procedures using mixed models that correct for these confounding factors have been developed and used in GWAS to minimize the detection of false positives (Yu et al. 2006; Kang et al. 2008). Similarly, some of the real associations fail to be detected due to lack of statistical power. For this reason, GWAS is widely considered as hypothesis generation step, and the marker-trait associations detected are validated through replication in independent association studies or linkage studies in bi-parental populations, to be considered for further applications. Considering this fact, our study was designed to detect SNPs that are significantly associated with kernel-Zn and Fe through GWAS, and these leads were validated in three independent bi-parental populations.

GWAS was performed using multiple statistical models, and the MLM correction for population structure and kinship was found to control the genomic inflation to the minimum level. Marker-trait associations were declared significant based on significance threshold corrected for multiple testing corrections taking average extent of genome-wide LD into consideration. For validation of SNPs, three DH populations were developed and phenotyped at two environments. We selected a higher number of SNPs for testing in the bi-parental populations by including the bottom 0.1 percentile of the distribution to test if SNPs had a significant effect on the phenotype in specific bi-parental populations.

In total 11 SNPs each for kernel-Zn and Fe ($P \leq 0.01$ and $R^2 \geq 5\%$) were found to have a significant effect on these traits in at least one population. However, it should be noted that about one-third of the SNPs that were selected for testing were not polymorphic in any of the parental combinations, limiting the ability of them being tested or validated in the present study. Some of the SNPs that were tested explained large proportion of phenotypic variance in individual bi-parental populations, though these could have been slightly over-estimated in single-marker QTL analysis. SNPs S7_173181688 and S7_173181689,

located at physical coordinates chr 7: 173,181,688 and chr 7: 173181689, respectively, were shown to have the largest proportion of variance explained for kernel-Zn in the bi-parental populations studied. These will be further tested in breeding populations for their usefulness in selecting lines with high Zn. Similarly, SNPs on chromosome 8, around 164 and 167 Mb were found to explain a large proportion of variance for kernel-Fe. Considering the SNPs that were tested to be significant in trait expression, approximately five genomic regions, represented by one to many SNPs were identified. A region on chromosome 7 within 1 Mb between 173 and 174 Mb was found to be having significant effect for both kernel-Zn and Fe, and will be closely followed in later studies towards using them as breeding targets in Zn and Fe improvement.

QTLs and candidate genes

Previous studies have reported QTL mapping and meta-QTL analysis for kernel-Zn and Fe in maize (Lung'aho et al. 2011; Qin et al. 2012; Šimić et al. 2012; Baxter et al. 2013; Jin et al. 2013). We compared the genomic positions of these QTLs against the ones detected in this study to determine if any of these fall within reported QTL intervals (Table 4). For kernel-Zn, reported chromosomal bins 3.04 (Qin et al. 2012), 4.06, 5.04, (Jin et al. 2013) and 9.06–07 (Qin et al. 2012; Jin et al. 2013) were found to have significant SNPs for kernel-Zn in this study. Similarly, for kernel-Fe, chromosomal bins, 2.04–07 (Qin et al. 2012; Jin et al. 2013), 3.04–06, 4.06 (Jin et al. 2013), 5.01 (Lung'aho et al. 2011; Baxter et al. 2013) and 8.06 (Šimić et al. 2012) were found to have significant SNPs for kernel-Fe detected in this study. There has been conflicting reports on identifying co-localized QTLs for the two traits in accordance to the phenotypic correlation between kernel-Zn and Fe (Qin et al. 2012; Šimić et al. 2012; Jin et al. 2013, 2015). In our study, we have observed only limited co-localization of the genomic regions controlling these two traits, like the ones on chromosome 4 (161–167 Mb), 7 (173–174 Mb) and 9 (136 Mb).

Some of the marker-trait associations identified in this study were co-located within genes which were previously reported to be linked to Zn and Fe uptake, transport or localization in plants. Among the significant associations detected, only one gene (different SNPs from gene GRMZM2G489070 on chromosome 9) was found to be common for both kernel-Zn and Fe. The Zn-associated SNP S8_80619983 near GRMZM2G311974 gene model possesses No Apical Meristem (NAC) domain transcriptional regulator super family protein. Molecular studies have shown that NAC family transcription factors regulates Fe and Zn remobilization from source organs to developing seeds associated with senescence (Ricachenevsky et al. 2013). A NAC transcription factor was also found to increase grain Fe and Zn content

in wheat (Uauy et al. 2006). One of the SNPs associated with kernel-Fe (S5_5104719) was located within the gene model GRMZM2G016756, which has active domains of the transcription factor, phytochrome-interacting factor-4, known to regulate auxin biosynthesis (Franklin et al. 2011). Auxin plays a role in the root morphology in response to Fe availability (Chen et al. 2010; Shen et al. 2015). A recent study has also demonstrated evidence of cross-talk between Zn homeostasis and auxin in Arabidopsis (Rai et al. 2015). Within the gene model GRMZM2G147698, two Fe associated SNPs, S1_64238426 and S1_64238509 were located, and this gene has a myb-like DNA-binding domain that is associated with Fe and Zn transport during nutrient deficiency (Shen et al. 2008). On Chromosome 1, SNPs S1_81549746 and S1_81549744 associated with kernel-Fe are within the gene model GRMZM2G302373, which had glutamine-*s* transferase activity that is involved in stress responses including heavy-metal toxicity and synthesis of phenolic compounds including cinnamic acid (Dixon et al. 2002). S4_167189737 associated with kernel-Fe was within the GRMZM2G168369 gene model related to zinc finger C3HC4 TYPE (RING FINGER) family protein. This gene encodes a cysteine-rich domain of 40–60 residues that coordinates two Zn ions and play a key role in the ubiquitination pathway (Lorick et al. 1999). This family of proteins are weakly up-regulated by Fe deficiency in Arabidopsis, giving an indication that they are linked with Fe homeostasis (Buckhout et al. 2009).

In conclusion, the present study is the first report of a Genome-wide association study (GWAS) using high-density genomic data conducted for detecting marker-trait associations for kernel-Zn and Fe in maize. The study identified about 20 and 26 SNPs, respectively, for kernel-Zn and Fe, respectively. A subset of the marker-trait associations was validated using single marker QTL analysis in three biparental populations. Whereas some of the genomic regions identified in this study were novel, others were located in already reported QTL intervals. Some of the identified SNPs were located within many known genes involved in uptake, transport and localization of Fe and Zn in plants. More studies are being carried out to validate the utility of the markers identified in this study in the breeding lines and populations, as a precursor to marker-based breeding for bio-fortification of tropical maize for increased kernel-Zn and Fe contents.

Acknowledgements This research was supported by HarvestPlus (<http://www.harvestplus.org>) and MAIZE CRP. We thank German Minngram, Jose Luis Coyac and Andrea Cruz for managing the trials, and Aldo Rosales and Ruben Basilio for carrying out the micronutrient analyses. We also thank the double-haploid team. The authors also thank Gordon Huestis for facilitating DNA extraction and technical help provided by Alberto Vergara at CIMMYT, Mexico. We gratefully acknowledge the technical inputs provided by Mr. Ashrafal Alam at CIMMYT, Hyderabad. The authors also thank our colleagues Jill Cairns, Kevin Pixley and Dan Makumbi for the information provided

regarding additional traits of the top high zinc germplasm. We thank CIMMYT colleagues, Michael Olsen, Xueaci Zhang and Jill Cairns for reviewing the manuscript. We acknowledge the Department of Biotechnology (DBT) for providing the financial support to the first author.

Author contribution statement NPR, RB, SKN and VH designed the experiments. NPR coordinated the field trials and phenotyping. RB and NPR coordinated the sample collection, DNA extraction and genotyping. VH, SKN, RB, NPR, WBS, ZR and GRS analyzed the data. VH, SKN and RB wrote the manuscript and all the authors revised and edited it.

Compliance with ethical standards

The authors declare that the experiments comply with the laws of Mexico.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ahmadi M, Wiebold WJ, Beuerlein JE, Eckert DJ, Schoper J (1993) Agronomic practices that affect corn kernel characteristics. *Agron J* 85:615–619
- Alvarado M, López M, Vargas A et al (2015) META-R (multi environment trial analysis with R for windows) Version 4.1, pp 6–8 (2015). <http://hdl.handle.net/11529/10201>
- Anuradha K, Agarwal S, Batchu AK et al (2012) Evaluating rice germplasm for iron and zinc concentration in brown rice and seed dimensions. *Plant Breed Rev* 4:19–25
- Atlin GN, Palacios N, Babu R et al (2010) Quality protein maize: progress and prospects. *Plant Breed Rev*. <https://doi.org/10.1002/9780470880579.ch3>
- Babu R, Rojas NP, Gao S et al (2013) Validation of the effects of molecular marker polymorphisms in LcyE and CrtRB1 on provitamin A concentrations for 26 tropical maize populations. *Theor Appl Genet* 126:389–399. <https://doi.org/10.1007/s00122-012-1987-3>
- Banziger M, Long J (2000) The potential for increasing the iron and zinc density of maize through plant-breeding. *Food Nutr Bull* 21:397–400
- Banziger M, Edmeades GO, Beck D, Bellon M (2000) Breeding for drought and nitrogen stress tolerance in maize: from theory to practice. CIMMYT, Mexico, DF
- Bashir K, Ishimaru Y, Nishizawa NK (2012) Molecular mechanisms of zinc uptake and translocation in rice. *Plant Soil* 361:189–201. <https://doi.org/10.1007/s11104-012-1240-5>
- Baxter IR, Gustin JL, Settles AM, Hoekenga OA (2013) Ionic characterization of maize kernels in the intermated b73 × mo17 population. *Crop Sci* 53:208–220. <https://doi.org/10.2135/cropsci2012.02.0135>
- Bishop DT, Williamson JA (1990) The power of identity-by-state methods for linkage analysis. *Am J Hum Genet* 46:254–265

- Bouis HE (2000) Enrichment of food staples through plant breeding: a new strategy for fighting micronutrient malnutrition. *Nutrition* 16:701–704. [https://doi.org/10.1016/S0899-9007\(00\)00266-5](https://doi.org/10.1016/S0899-9007(00)00266-5)
- Bouis HE, Welch RM (2010) Biofortification-A sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Sci* 50:S-20–S-32. <https://doi.org/10.2135/cropsci2009.09.0531>
- Brkić I, Šimić D, Zdunić Z et al (2004) Genotypic variability of micronutrient element concentrations in maize kernels. *Cereal Res Commun* 32:107–112
- Buckhout TJ, Yang TJW, Schmidt W (2009) Early iron-deficiency-induced transcriptional changes in Arabidopsis roots as revealed by microarray analyses. *BMC Genom* 10:147. <https://doi.org/10.1186/1471-2164-10-147>
- Cairns JE, Cossa J, Zaidi PH et al (2013) Identification of drought, heat, and combined drought and heat tolerant donors in maize. *Crop Sci*. <https://doi.org/10.2135/cropsci2012.09.0545>
- Cakmak I (2008) Enrichment of cereal grains with zinc: agronomic or genetic biofortification? *Plant Soil* 302:1–17. <https://doi.org/10.1007/s11104-007-9466-3>
- Chakraborti M, Prasanna BM, Hossain F, Singh AM, Guleria SK (2009) Genetic evaluation of kernel Fe and Zn concentrations and yield performance of selected Maize (*Zea mays* L.) genotypes. *Range Manag Agrofor* 30:109–114
- Chakraborti M, Prasanna BM, Hossain F (2011) Identification of kernel iron- and zinc-rich maize inbreds and analysis of genetic diversity using microsatellite markers. *J Plant Biochem Biotechnol* 20:224–233. <https://doi.org/10.1007/s13562-011-0050-9>
- Chen WW, Yang JL, Qin C et al (2010) Nitric oxide acts downstream of auxin to trigger root ferric-chelate reductase activity in response to iron deficiency in Arabidopsis. *Plant Physiol* 154:810–819. <https://doi.org/10.1104/pp.1104.161109>
- Chomba E, Westcott CM, Westcott JE et al (2015) Zinc absorption from biofortified maize meets the requirements of young rural zambian children. *J Nutr* 145:514–519. <https://doi.org/10.3945/jn.114.204933>
- Cimmyt (2005) Laboratory protocols: CIMMYT applied molecular genetics laboratory, 3rd edn. CIMMYT, Mexico, DF
- Core Team R (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Cui Z, Luo J, Qi C et al (2016) Genome-wide association study (GWAS) reveals the genetic architecture of four husk traits in maize. *BMC Genom* 17:946. <https://doi.org/10.1186/s12864-016-3229-6>
- Diepenbrock CH, Gore MA (2015) Closing the divide between human nutrition and plant breeding. *Crop Sci* 55:1437–1448. <https://doi.org/10.2135/cropsci2014.08.0555>
- Diez-Altare C, Bornemiza E (1967) The localization of ZINC-65 in Germinating corn tissues. *Plant Soil* 26:175–188
- Dixon DP, Laphorn A, Edwards R (2002) Plant glutathione transferase. *Genome Biol* 3:3-3004.1–3-3004.10
- Eckardt NA (2000) Sequencing the rice genome. *Plant Cell* 12:2011–2017. <https://doi.org/10.1105/tpc.12.11.2011>
- Eide D, Broderius M, Fett J, Guerinot ML (1996) A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc Natl Acad Sci USA* 93:5624–5628. <https://doi.org/10.1073/pnas.93.11.5624>
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Franklin KA, Lee SH, Patel D et al (2011) Phytochrome-interacting factor 4 (PIF4) regulates auxin biosynthesis at high temperature. *PNAS* 108:20231–20235. <https://doi.org/10.1073/pnas.11110682108>
- Galicia L, Rosales A, Gutierrez G et al (2012) Manual de laboratorio de calidad nutricional de maíz. CIMMYT, Mexico
- Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. *Curr Opin Plant Biol* 13:167–173. <https://doi.org/10.1016/j.pbi.2010.01.004>
- Graham RD, Welch RM (1996) Breeding for staple food crops with high micronutrient density. *Int Food Policy Res Inst* 3:79
- Gross J, Stein RJ, Fett-Neto AG et al (2003) Iron homeostasis related genes in rice. *Genet Mol Biol* 26:477–497. <https://doi.org/10.1590/S1415-47572003000400012>
- Grotz N, Fox T, Connolly E et al (1998) Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency. *PNAS* 95:7220–7224. <https://doi.org/10.1073/pnas.95.12.7220>
- Guerrant RL, Lima AA, Davidson F (2000) Micronutrients and infection: interactions and implications with enteric and other infections and future priorities. *JID* 182:134–138. <https://doi.org/10.1086/315924>
- Gupta SK, Velu G, Rai KN, Sumalini K (2009) Association of grain iron and zinc content with grain yield and other traits in pearl millet. *Crop Improv* 36(2):4–7
- Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. *Theor App Genet* 38:226–231. <https://doi.org/10.1007/BF01245622>
- Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol* 65:531–551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Jin T, Zhou J, Chen J et al (2013) The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. *Breed Sci* 63:317–324. <https://doi.org/10.1270/jsbbs.63.317>
- Jin T, Chen J, Zhu L et al (2015) Comparative mapping combined with homology-based cloning of the rice genome reveals candidate genes for grain zinc and iron concentration in maize. *BMC Genet*. <https://doi.org/10.1186/s12863-015-0176-1>
- Kabir KA, Haque M, Hossain MA et al (2003) Breeding for iron-dense rice in Bangladesh. In: Mew TW, Brar DS, Peng S et al (eds) Rice science: innovations and impact for livelihood. Metro Manila, pp 397–402
- Kang HM, Zaitlen NA, Wade CM et al (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>
- Kapil U, Bhavna A (2002) Adverse effects of poor micronutrient status during childhood and adolescence. *Nutr Rev* 60:S84–S90. <https://doi.org/10.1301/00296640260130803>
- Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. *Annu Rev Plant Biol* 63:131–152. <https://doi.org/10.1146/annurev-arplant-042811-105522>
- Kumar AA, Reddy BVS, Ramaiah B et al (2009) Genetic variability and plant character association of grain Fe and Zn in selected core collection accessions of sorghum germplasm and breeding lines. *J SAT Agric Res* 7:1–4
- Lee S, Jeong HJ, Kim SA et al (2010) OsZIP5 is a plasma membrane zinc transporter in rice. *Plant Mol Biol* 73:507–517. <https://doi.org/10.1007/s11103-010-9637-0>
- Lorick KL, Jensen JP, Fang S et al (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci USA* 96:11364–11369. <https://doi.org/10.1073/pnas.96.20.11364>
- Lu Y, Shah T, Hao Z et al (2011) Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. *PLoS One*. <https://doi.org/10.1371/journal.pone.0024861>

- Lung'aho MG, Mwaniki AM, Szalma SJ et al (2011) Genetic and physiological analysis of iron biofortification in maize kernels. *PLoS One* 6:1–10. <https://doi.org/10.1371/journal.pone.0020429>
- Maziya-Dixon B, Kling JG, Menkir A, Dixon A (2000) Genetic variation in total carotene, iron, and zinc contents of maize and cassava genotypes. *Food Nutr Bull* 21:419–422. <https://doi.org/10.1177/156482650002100415>
- Menkir A (2008) Genetic variation for grain mineral content in tropical-adapted maize inbred lines. *Food Chem* 110:454–464. <https://doi.org/10.1016/j.foodchem.2008.02.025>
- Obata T, Witt S, Liseč J et al (2015) Metabolite profiles of maize leaves in drought, heat and combined stress field trials reveal the relationship between metabolism and grain yield. *Plant Physiol* 169:2665–2683. <https://doi.org/10.1104/pp.15.01164>
- Ortiz-Monasterio I, Palacios-Rojas N, Meng E et al (2007) Enhancing the mineral and vitamin content of wheat and maize through plant breeding. *J Cereal Sci* 46:293–307
- Panda SK, Baluska F, Matsumoto H (2009) Aluminum stress signaling in plants. *Plant Signal Behav* 4:592–597. <https://doi.org/10.4161/psb.4.7.8903>
- Pixley K (2013) Biofortification of maize with provitamin A carotenoids. In: Tanumihardjo S (ed) Carotenoids and human health. Nutrition and health. Humana Press, Totowa. <https://doi.org/10.1007/978-1-62703-203-2>
- Prasanna BM, Mazumdar S, Chakraborti M et al (2011) Genetic variability and genotype x environment interactions for kernel iron and zinc concentrations in maize (*Zea mays*) genotypes. *Indian J Agric Sci* 81:704–711
- Price AL, Patterson NJ, Plenge RM et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38:904–909. <https://doi.org/10.1038/ng1847>
- Qin H, Cai Y, Liu Z et al (2012) Identification of QTL for zinc and iron concentration in maize kernel and cob. *Euphytica* 187:345–358. <https://doi.org/10.1007/s10681-012-0692-2>
- Rai V, Sanagala R, Sinilal B et al (2015) Iron availability affects phosphate deficiency-mediated responses, and evidence of cross-talk with auxin and zinc in Arabidopsis. *Plant Cell Physiol* 56:1107–1123. <https://doi.org/10.1093/pcp/pcv035>
- Rashid Z, Singh PK, Vemuri H, Zaidi PH, Prasanna BM, Nair SK (2018) Genome-wide association study in Asia-adapted tropical maize reveals novel and explored genomic regions for sorghum downy mildew resistance. *Sci Rep* 8:366. <https://doi.org/10.1038/s41598-017-18690-3>
- Remington DL, Thornsberry JM, Matsuoka Y et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *PNAS* 98:11479–11484. <https://doi.org/10.1073/pnas.201394398>
- Ricachenevsky FK, Menguer PK, Sperotto RA (2013) kNACKing on heaven's door: how important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds? *Front Plant Sci* 4:1–7. <https://doi.org/10.3389/fpls.2013.00226>
- Romay MC, Millard MJ, Glaubitz JC et al (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol* 14:R55. <https://doi.org/10.1186/gb-2013-14-6-r55>
- Sehgal D, Autrique E, Singh R et al (2017) Identification of genomic regions for grain yield and yield stability and their epistatic interactions. *Sci Rep* 7:41578. <https://doi.org/10.1038/srep41578>
- Sharma A, Chauhan RS (2008) Identification of candidate gene-based markers (SNPs and SSRs) in the zinc and iron transporter sequences of maize (*Zea mays* L.). *Curr Sci* 95:8–25
- Shen J, Xu X, Li T et al (2008) An MYB transcription factor from *Malus xiaojinensis* has a potential role in iron. *Nutrition* 50:1300–1306. <https://doi.org/10.1111/j.1744-7909.2008.00761.x>
- Shen C, Yue R, Sun T et al (2015) OsARF16, a transcription factor regulating auxin redistribution, is required for iron deficiency response in rice (*Oryza sativa* L.). *Plant Sci* 231:148–158. <https://doi.org/10.1016/j.plantsci.2014.12.003>
- Šimić D, Mladenović Drinić S, Zdunić Z et al (2012) Quantitative trait loci for biofortification traits in maize grain. *J Hered* 103:47–54. <https://doi.org/10.1093/jhered/esr122>
- Smale M, Byerlee D, Jayne T (2013) Maize revolutions in sub-Saharan Africa. In: Otsuka K, Larson DF (eds) An African green revolution: finding ways to boost productivity on small farms. The International Bank for Reconstruction and Development/The World Bank, Dordrecht, The Netherlands, p 165–195
- Suwarno WB, Pixley KV, Palacios-Rojas N et al (2014) Formation of heterotic groups and understanding genetic effects in a provitamin A biofortified maize breeding program. *Crop Sci* 54(1):14–24. <https://doi.org/10.2135/cropsci2013.02.0096>
- Tiwari C, Wallwork H, Arun B et al (2016) Molecular mapping of quantitative trait loci for zinc, iron and protein content in the grains of hexaploid wheat. *Euphytica* 207:563–570. <https://doi.org/10.1007/s10681-015-1544-7>
- Uauy C, Distelfeld A, Fahima T et al (2006) A NAC gene regulating senescence improves grain protein, zinc, and iron content in wheat. *Science* 314:1298–1301. <https://doi.org/10.1126/science.1133649>
- Velu G, Bhattacharjee R, Rai KN et al (2008) A simple and rapid screening method for grain zinc content in pearl millet. *SAT* eJ 6:2006–2009
- Velu G, Singh R, Peña J (2011) Breeding for enhanced zinc and iron concentration in CIMMYT spring wheat germplasm. *Czech J Genet Plant Breed* 47:S174–S177
- Vert G, Briat JF, Curie C (2001) Arabidopsis IRT2 gene encodes a root-periphery iron transporter. *Plant J* 26:181–189. <https://doi.org/10.1046/j.1365-3113X.2001.01018.x>
- Warburton ML, Xianchun X, Crossa J et al (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci* 42:1832–1840. <https://doi.org/10.2135/cropsci2002.1832>
- Waters BM (2002) Characterization of FRO1, a pea ferric-chelate reductase involved in root iron acquisition. *Plant Physiol* 129:85–94. <https://doi.org/10.1104/pp.010829>
- Welch RM, Graham RD (2002) Breeding crops for enhanced micronutrient content. *Plant Soil* 245:205–214. <https://doi.org/10.1023/A:1020668100330>
- Witt S, Galicia L, Liseč J et al (2012) Metabolic and phenotypic responses of greenhouse-grown maize hybrids to experimentally controlled drought stress. *Mol Plant* 5:401–417. <https://doi.org/10.1093/mp/ssr102>
- Yan J, Shah T, Warburton ML et al (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One*. <https://doi.org/10.1371/journal.pone.0008451>
- Yang W, Guo Z, Huang C et al (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat Commun* 5:5087. <https://doi.org/10.1038/ncomms6087>
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160. <https://doi.org/10.1016/j.copbio.2006.02.003>
- Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38:203–208. <https://doi.org/10.1038/ng1702>
- Yu J, Zhang Z, Zhu C et al (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. *Plant Genome* J 2:63. <https://doi.org/10.3835/plantgenome2008.09.0009>

Zhang Z, Wei T, Zhong Y, Li X, Huang J (2016) Construction of a high-density genetic map of *Ziziphus jujuba*. *Tree Genet Genomes* 4:1–10. <https://doi.org/10.1007/s11295-016-1032-9>

Zhao H, Eide D (1996) The ZRT2 gene encodes the low affinity zinc transporter in *Saccharomyces cerevisiae*. *J Biol Chem* 271:23203–23210. <https://doi.org/10.1074/jbc.271.38.23203>

Affiliations

Vemuri Hindu^{1,3} · Natalia Palacios-Rojas² · Raman Babu^{1,4} · Willy B. Suwarno^{2,5} · Zerka Rashid¹ · Rayalcheruvu Usha³ · Gajanan R Saykhedkar^{1,6} · Sudha K. Nair¹

Vemuri Hindu
hinduvemuri@gmail.com

Natalia Palacios-Rojas
n.palacios@cgiar.org

Raman Babu
raman.babu@pioneer.com

Willy B. Suwarno
bayuardi@gmail.com

Zerka Rashid
m.zerka@cgiar.org

Rayalcheruvu Usha
ushatirupathi@gmail.com

Gajanan R Saykhedkar
saigajanan33@gmail.com

¹ Asia Regional Maize Program, International Maize and Wheat Improvement Center (CIMMYT), ICRISAT Campus, Patancheru, Hyderabad, Telangana 502324, India

² International Maize and Wheat Improvement Center (CIMMYT), Km 45 Carretera Mexico-Veracruz, 56130 Texcoco, Mexico

³ Sri Padmavati Mahila Visvavidyalayam (Women's University), Tirupati, Andhra Pradesh 517502, India

⁴ Present Address: Multi-Crop Research Center (MCRC), DuPont Pioneer, Hyderabad, Telangana 500078, India

⁵ Present Address: Department of Agronomy and Horticulture, Faculty of Agriculture, Bogor Agricultural University, Jl. Meranti Kampus IPB Dramaga, Bogor 16680, Indonesia

⁶ Present Address: Project Director, SPMESM, Dr. Hedgewar Hospital, Aurangabad, Maharashtra 431005, India