



OPEN

## Genome-wide association study and expression of candidate genes for Fe and Zn concentration in sorghum grains

Niranjan Ravindra Thakur<sup>1,2</sup>, Sunita Gorthy<sup>1</sup>, Anil Kumar Vemula<sup>1</sup>, Damaris A. Odeny<sup>1</sup>, Pradeep Ruperao<sup>1</sup>, Pramod Ramchandra Sargar<sup>1,2</sup>, Shivaji Pandurang Mehtre<sup>2</sup>, Hirakant V. Kalpande<sup>2</sup> & Ephrem Habyarimana<sup>1</sup>✉

Sorghum germplasm showed grain Fe and Zn genetic variability, but a few varieties were biofortified with these minerals. This work contributes to narrowing this gap. Fe and Zn concentrations along with 55,068 high-quality GBS SNP data from 140 sorghum accessions were used in this study. Both micronutrients exhibited good variability with respective ranges of 22.09–52.55 ppm and 17.92–43.16 ppm. Significant marker-trait associations were identified on chromosomes 1, 3, and 5. Two major effect SNPs (S01\_72265728 and S05\_58213541) explained 35% and 32% of Fe and Zn phenotypic variance, respectively. The SNP S01\_72265728 was identified in the cytochrome P450 gene and showed a positive effect on Fe accumulation in the kernel, while S05\_58213541 was intergenic near Sobic.005G134800 (zinc-binding ribosomal protein) and showed negative effect on Zn. Tissue-specific *in silico* expression analysis resulted in higher levels of Sobic.003G350800 gene product in several tissues such as leaf, root, flower, panicle, and stem. Sobic.005G188300 and Sobic.001G463800 were expressed moderately at grain maturity and anthesis in leaf, root, panicle, and seed tissues. The candidate genes expressed in leaves, stems, and grains will be targeted to improve grain and stover quality. The haplotypes identified will be useful in forward genetics breeding.

After a decades-long fall and five years of stability since 2014, the global Prevalence of Undernourishment (PoU) spiked during 2019 and 2020 due to the COVID-19 pandemic<sup>1</sup>. In 2020, 21% of the population in Africa was hungry<sup>1</sup>. Asia's enormous population houses 54% of the world's hungry<sup>1</sup>. According to the World Health Organization<sup>2–4</sup>, 149.2 million children under five were stunted, with diet-related deaths amounting to 45%, whereas 462 million adults were underweighting in 2014. African region tops this critical situation with 56.2 million children, followed by 49.8 million in South-East Asia, and 22.9 million children in the Eastern Mediterranean region<sup>5</sup>. Malnutrition is the inadequate consumption of micronutrients like iodine, vitamin A, iron, and zinc which affects worldwide public health<sup>2,6</sup>. Therefore, inadequate nutrition threatens the health and development of populations worldwide, especially children and pregnant women in countries with lower income.

The importance of iron (Fe) in human food cannot be overemphasized. Iron deficiency anemia is a common type of anemia described as blood lacking adequate healthy red blood cells carrying oxygen to the body's tissues<sup>7–9</sup>. Anemia decreases physical and mental capabilities but is rarely discovered, and when worsens it results in death. Anemia during pregnancy can impede growth and development due to preterm delivery, low-birth-weight infants, maternal death, and reduced iron storage. In 2019, 29.9% of reproductive-age women (nearly half a billion women aged 15–49), 36.5% of pregnant women, and 39.8% of 6–59-month-olds (269 million children) had anemia<sup>10</sup>.

On the other hand, zinc, one of the most abundant trace elements, is the most widely distributed trace element in the body after iron. Zinc has a structural role in over 2500 transcription factors, is necessary for the activity of over 300 proteins structure and enzymes activity, and controls thousands of genes and gene expressions<sup>11,12</sup>. Zinc is important in DNA synthesis, cell proliferation, protein synthesis, wound healing, and immune system support. It considerably impacts macrophages, neutrophils, and other complementary activity at the microcellular level<sup>13–15</sup>. Two billion people are still at risk of zinc deficiency, with infants, toddlers, and pregnant and

<sup>1</sup>International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Telangana, India. <sup>2</sup>Vasantrao Naik Marathwada Agriculture University, Parbhani, Maharashtra, India. ✉email: Ephrem.Habyarimana@icrisat.org

lactating women being the most vulnerable due to their higher zinc requirements<sup>16,17</sup>. Zinc must be consumed or supplemented because the body cannot store it; its insufficiency is a global health issue, especially in developing nations like India, Pakistan, Ethiopia, Indonesia, and Vietnam<sup>18</sup>.

External nutrient supplementation and fortification are the most widely used methods to counter micronutrient malnutrition. Biofortified foods can also be produced through plant breeding and/or agronomic practices designed to increase the density of vitamins and minerals in the crop of interest. It is now acknowledged that developing high-nutrient cultivars is cheap, and making it affordable for low-income populations is a promising strategy to reduce micronutrient malnutrition.

Sorghum is an ideal crop under current and future climate change scenarios because it generates a good net return with low input and can feed a larger population, especially in semi-arid ecologies where wheat (*Triticum aestivum* L.), maize (*Zea mays* L.), and rice (*Oryza sativa*) cannot be produced sustainably due to climate adversities, particularly high temperatures and increased drought stress<sup>19,20</sup>. Sorghum is a staple food for more than 500 million people in Africa and Asia<sup>21</sup>. The estimated global sorghum consumption is expected to reach 26.5 million metric tons by the year 2026. Nigeria emerged as the leading consumer of sorghum in 2021, consuming a staggering 5.7 million metric tons. India, China, and Ethiopia were next, with 3.9, 3.7, and 3.5 million metric tons, respectively<sup>22</sup>. Due to its low glycemic index and high antioxidant content, sorghum grain is a popular gluten-free and celiac-safe diet<sup>23</sup>, and its consumption is expected to increase in developed countries driven by not only its life-promoting properties but also by its cultivation as a substitute of major cereals in those parts of the world. Biofortifying sorghum using conventional plant breeding and molecular technologies approaches can improve sorghum as a nutrient-dense and climate-resilient crop.

Previous studies on grain Fe and Zn concentrations in sorghum found significant genetic variability and strong heritability<sup>24,25</sup>. Madhusudhana et al.<sup>26</sup> recently observed a highly significant G × E for Fe, but not for Zn. A significant G × E indicates that a genotype can perform well in particular environments but not others. Data from several trials showed high broad sense heritability for sorghum grain Fe (> 85%) and Zn concentrations (> 82%)<sup>27</sup>. Kumar et al.<sup>23</sup> concluded that additive gene action controls Zn, and therefore high Zn lines can be developed by increasing the frequency of favorable alleles through intercrosses and other population improvement strategies. Additive and non-additive gene actions were found to regulate Fe concentration<sup>23</sup>, implying that this trait can be improved through heterosis breeding on the one hand, and crossing and selecting superior progeny, on the other. The highly significant and positive correlation between general combining ability (GCA) and line performance per se suggested that parental genotype performance per se can predict the grain Fe and Zn concentrations in hybrids and breeding lines<sup>23</sup>.

Although sorghum germplasm has shown large variability and genetic heritability for iron and zinc content<sup>25</sup>, only a few Fe and Zn biofortified varieties have been released thus far<sup>28</sup>. Indeed, the first India's biofortified sorghum was released only two years ago, in 2022<sup>29</sup>. This work is, therefore, a contribution to the identification of major quantitative traits loci (QTLs) and genes governing Fe and Zn biofortification for use in breeding e.g., through introgressions in a sustained marker-assisted selection to create new biofortified varieties and improve the nutritional quality of the existing cultivars.

Recently, biparental mapping populations were used to identify QTLs for Fe and Zn in sorghum<sup>24</sup>. However, the QTLs found using this method were of low resolution. In contrast, linkage disequilibrium (LD)-based association mapping (Genome-Wide Association Study) can boost mapping resolution by representing a more diverse gene pool and accounting for past meiotic events. This technique was used in various previous research works to find markers non-randomly associated with the phenotype of interest in a broader process of forward genetics and breeding. Girma et al.<sup>30</sup>, Enyew et al.<sup>31</sup>, Wondimu et al.<sup>32</sup> carried out GWAS on sorghum agro-morphological traits, while other GWAS were successfully conducted on nutritional traits such as polyphenol<sup>33,34</sup>, protein, fat, and starch<sup>35</sup>, mineral traits<sup>36</sup>, and biomass related traits<sup>37</sup>. Several genomic regions governing grain Fe and Zn have already been identified in other cereals such as Pearl Millet<sup>38,39</sup>, Wheat<sup>40,41</sup>, Maize<sup>42</sup>, Barley<sup>43</sup>, and Rice<sup>44</sup>. However, no reports on identifying SNPs and genes controlling grain Fe and Zn in sorghum are available. This work was therefore undertaken to close this gap; it is the first to report the marker traits associations (MTAs) for grain Fe and Zn in sorghum using SNPs and minicore lines. We aimed at identifying SNPs associated with grain Fe and Zn concentration in the sorghum kernel, mining underlying candidate genes using a part of the ICRISAT's minicore set of sorghum collection, and implementing several GWAS models. ICRISAT's genebank is the repository of the largest sorghum germplasm collection i.e., over 42,880 accessions from over 94 countries worldwide (<http://genebank.icrisat.org/>). The mini core, a gateway to the germplasm, is highly genetically diverse and is an efficient option for carrying out association mapping and allele mining for traits of interest<sup>45</sup>. The outcome of this work is, therefore, expected to benefit the scientific community across the globe.

## Materials and methods

### Multi-environment field trials and plant materials

The experimental materials for the current study consisted of 140 diverse germplasm accessions representing a part of a sorghum minicore maintained at the International Crops Research for the Semi-Arid Tropics (ICRISAT, Patancheru), from 58 different countries (Supplementary Table S1). These germplasm accessions were evaluated for grain Fe and Zn content (ppm) in post-rainy seasons of two consecutive years across two locations. The seasons included post rainy 2020 (PR 2020) and post rainy 2021 (PR 2021), while the locations were the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT, Patancheru) (17.53°N, 78.27°E) located at an altitude of 545 m above mean sea level and Vasantrao Naik Marathwada Krishi Vidyapeeth (VNMKV, Parbhani) (18.45°N, 76.13°E) located at an altitude of 357 m above mean sea level. Soil testing analysis using the DTPA method for the experiment fields at Parbhani was done at the Department of Soil Science, COA, Parbhani, while, for ICRISAT fields it was done at Charles Renard Analytical Laboratory, ICRISAT, Patancheru. The available Fe

and Zn content in our research field was beyond the minimum required levels (4.50 to 6.50 mg kg<sup>-1</sup> Fe and 0.6 to 0.9 mg kg<sup>-1</sup> Zn)<sup>46</sup> and ample for normal growth and development of plants. Soil type at Parbhani was deep and black whereas, at ICRISAT it was shallow and light black. Irrigation was provided at each important crop stages to grow healthy plants including the grains development stage. The experiments were conducted in the post-rainy season to ensure the good quality of the seed. The experimental design was alpha lattice with two replications. Each germplasm was grown in a single row of 4 m long at both locations with inter and intra-row spacing of 60 × 15 cm. Standard agronomic practices were followed for successful crop development in each season.

### Phenotyping and estimation of Fe and Zn concentration in the grains

The panicles were harvested to measure the Fe and Zn content from the grains as they reached the physiological maturity stage. The panicles from five random plants from each plot were selfed before the flowering stage to obtain pure seed. Upon maturity, these panicles were harvested and stored separately in a dry cloth bag to produce clean grain samples for micronutrient analysis.

The harvested panicles underwent a 7-day sun drying to achieve a post-harvest grain moisture level of 10–12%, essential for mitigating fungal infections. Threshing was meticulously conducted in cloth bags to prevent sample contamination by extraneous particles. Subsequently, the seeds underwent cleaning to eliminate glumes, anthers, dust, and other contaminants. A composite seed sample weighing approximately 30 g was then collected for each plot. Analysis of grain Fe and Zn content was performed using X-ray fluorescence spectrometry (XRF) (Make: Hitachi High-Tech, Japan; model: X-Supreme8000), a calibrated, efficient, non-destructive, and cost-effective analytical method.

### Statistical analyses

A combined analysis of variance (ANOVA) was performed to assess the main and interaction effects of Environment (E) and Genotype (G), considering E, G, and replications as random effects. The individual variance of environments was modelled using the Residual Maximum Likelihood (REML) procedure using SAS Mixed procedure using SAS v9.4<sup>47</sup>. Best Linear Unbiased Predictors (BLUPs) were estimated using below equation:

$$Y_{ijkl} = M + E_i + R(E)_{j(i)} + G_k + GE_{jk} + e_{ijkl}$$

where  $Y_{ijkl}$ ,  $M$ ,  $E_i$ ,  $R(E)_{j(i)}$ ,  $G_k$ ,  $GE_{jk}$ , and  $e_{ijkl}$ , respectively, stand for the measurement on plot  $l$  in environment  $i$ , block  $j$ , containing genotype  $k$ , the overall mean of all plots in all environments, the effect of environment (trial)  $i$ , the effect of replicate  $j$  within environment  $i$ , the effect of genotype  $k$ , the interaction of genotype  $k$  with environment  $i$ , the plot residual.

The heritability was estimated as repeatability<sup>48</sup> using the formula:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_{gxe}^2/E) + (\sigma_e^2/r * E)}$$

where  $\sigma_g^2$ ,  $\sigma_{gxe}^2$ , and  $\sigma_e^2$  are genetic variance component, genotype × environment variance component and residual variance respectively;  $E$  and  $r$  are the number of environments and replications, respectively. The four best combinations of environments were selected for phenotypic data processing in this study, based upon the relatively high repeatability (Table 1) from the combined analysis of variance. All combinations of the environments are presented in Supplementary Table S2.

### Isolation of DNA and genotyping

Genomic DNA from the single plant of sorghum minicore samples was isolated from 30 days old seedlings using the QIAGEN DNAeasy 96 plant kit. Purity and quantity of the extracted DNA was determined using gel electrophoresis and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) respectively and finally diluted to 30 ng/μl. Genotype-by-sequencing (GBS) libraries were prepared using the restriction enzyme ApeK1 according to Elshire et al.<sup>49</sup>. SNPs were called using the TASSEL v5.2 GBS pipeline<sup>50</sup> against the *S. bicolor* BTx623 reference

Combination No	Environments*				Repeatability (%)	
	ICRISAT_20	ICRISAT_21	Parbhani_20	Parbhani_21	Fe	Zn
1	✓	✓	✓	✓	0.46	0.66
2		✓	✓	✓	0.41	0.57
3	✓		✓		0.45	0.64
4			✓	✓	0.56	0.74

**Table 1.** Repeatability-based combinations of environments to generate BLUPs for downstream GWAS analyses. \*ICRISAT\_20—post-rainy season of the year 2020 at International Crops Research Institute for the Semi-Arid Tropics, Patancheru; ICRISAT\_21—post-rainy season of the year 2021 at International Crops Research Institute for the Semi-Arid Tropics, Patancheru; Parbhani\_20—post-rainy season of the year 2020 at Vasantrao Naik Marathwada Agriculture University, Parbhani; Parbhani\_21—post-rainy season of the year 2021 at Vasantrao Naik Marathwada Agriculture University, Parbhani; Fe—iron; Zn—zinc.

genome v3.1.1 ([www.phytozome.net](http://www.phytozome.net))<sup>51</sup> using Bowtie v2.5.1<sup>52</sup> with default parameters. Raw SNPs were called from the 140 genotypes and subsequently filtered to retain bi-allelic markers with < 40% missing data as suggested in Ali et al.<sup>53</sup>, maximum of 20% of heterozygosity and a minimum of 5% minor allele frequency (MAF). The retained high-quality SNPs were used in this work for genetic diversity, population structure and marker-trait association analyses (MTAs). The MTAs were performed using R GAPIT 3 pipeline<sup>54</sup>.

### Population structure

The high-quality SNPs were used to determine the genetic distance between the sorghum accessions using R "amap" and "labdsv" packages. The phylogeny analysis was performed with the Euclidean method with 1000 bootstrap replications with R "ape" package<sup>55</sup> and the neighbor-joining tree was visualized in iTOL tree viewer<sup>56</sup>. The Principal Coordinates Analysis (PCo) between the accessions was measured with R "labdsv" package (<https://CRAN.R-project.org/package=labdsv>). The admixture analysis was performed with ADMIXTURE 1.3.0<sup>57</sup> with expected K in the range of 2 to 15 and the K with the lowest cross-validation error considered as optimal sub-populations. The genome-wide Linkage Disequilibrium (LD) was generated using the  $r^2$  values calculated with TASSEL v5.2<sup>50</sup>.

### Association mapping

We performed GWAS with multiple models viz., MLMM, SUPER, BLINK and FarmCPU. Manhattan and Quantile–Quantile (Q–Q) plots were visualized in RStudio<sup>58</sup> using GAPIT 3 package<sup>54</sup>. The spurious associations in GWAS were corrected using Bonferroni Correction (5% level of significance)<sup>59</sup> and significant MTAs, corresponding to putative QTLs for the studied traits, were determined by the *P-value* less than 0.05/m, with m being the number of markers<sup>60</sup>. In the present investigation, Bonferroni correction value is calculated at  $9.07 \times 10^{-5}$ . Further, the percentage of phenotypic variance explained (PVE) by all significant SNPs was generated either by GAPIT 3 built-in algorithms or calculated in RStudio as the squared correlation between the phenotype (BLUPs) and genotype of the SNP accounting for the above-described populations of environments. The Linux bash and *in-house* developed script was used to find the haplotype blocks and to identify the promising germplasms exhibiting high Fe and Zn content from the sorghum population.

### Candidate gene identification and *in silico* gene expression analysis

Single nucleotide polymorphisms (SNPs) explaining more than 7.5% of the phenotypic variance of their associated traits were identified and their genomic regions were further analyzed in the process of functional GWAS and candidate gene identification. To perform functional GWAS, an interval of 51.77 Kb upstream and downstream the SNP position was considered, based on a genome-wide linkage disequilibrium (LD) decay cut-off at  $r^2 = 0.1$ . Annotation details for genes and respective Gene Ontologies (GO) within each region were retrieved using the PhytoMine interface implemented in Phytozome<sup>61</sup>.

JGI Plant Gene Atlas (<https://plantgenetlas.jgi.doe.gov/>)<sup>62</sup> was used to check the tissue-specific expression of candidate genes and to access the expression data for 36 tissues at juvenile, vegetative, and reproductive stages. Gene expression was in the unit of Fragments per kilobase of transcripts per million mapped reads (FPKM). The expression heatmap was generated using TBtools II<sup>63</sup>.

### Ethics statement

This research does not involve the ethics of human and animal experiments.

### Institutional, national, and international guidelines and legislation statement

The seeds of sorghum used in this work were collected from ICRISAT's Genebank (<http://genebank.icrisat.org/>). The authors confirm that all methods applied in this work were performed in accordance with the relevant guidelines/regulations/legislation.

## Results

### Statistical analyses

The repeatability i.e., test-retest reliability, resulted relatively highest (56%) for Fe in combination 4, followed by combinations 1, 3 and 2 with 46, 45 and 41%, respectively. Similarly for Zn the highest repeatability values (74%) were observed in combination 4 followed by combinations 1, 3 and 2 with 66, 64 and 57%, respectively, as shown in Table 1.

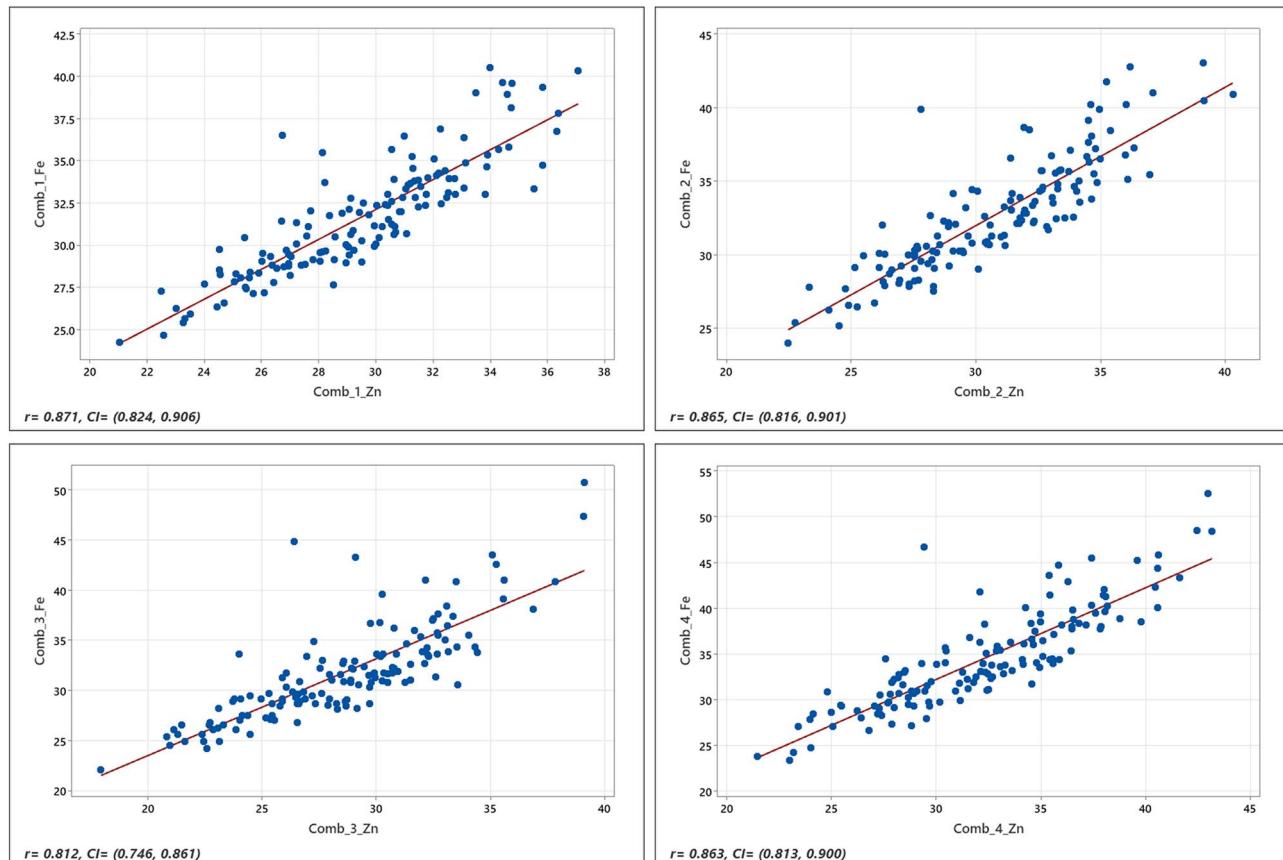
The combined ANOVA showed significant differences among the genotypes and a significant genotype  $\times$  environment for both Fe and Zn traits (Table 2). As depicted in Table 2 and in Fig. 1, the concentration of Fe and Zn varied over a wide range, and the correlation between the two micronutrients was high ( $r > 0.80$ ). The frequency distributions of the phenotypic data followed a normal distribution, as suggested by Shapiro–Wilks normality test<sup>64</sup>; and Ryan-Joiner statistics<sup>65</sup> (Fig. 2).

### Genetic diversity, population structure, and linkage disequilibrium (LD)

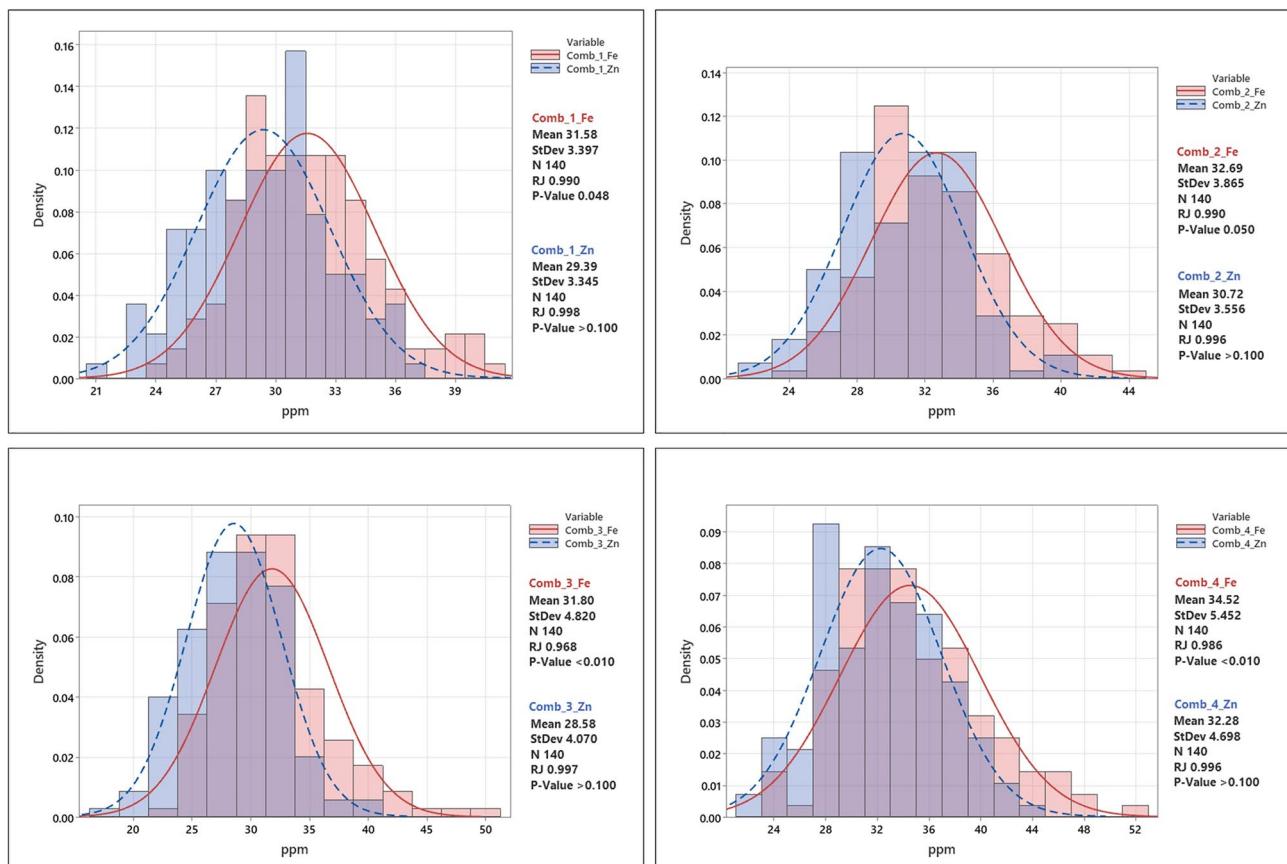
We retained 55,068 high quality SNPs that were well distributed across the genome (Fig. 3) with a density of ~ 5505 markers per chromosome (Chr) and around 80 markers per Mb region. Genomic regions with high marker density were observed on Chr 01 and Chr 02 with an average magnitude of 91 markers per Mb on both chromosomes. Chr 06 has averaged 86 markers per 1 Mb region with high-density regions between 47 and 52 Mb. Chr 08 and 09 also showed high-density regions between 57–63 Mb and 55–74 Mb, respectively. Linkage-disequilibrium decay (LD decay) is determined using the entire set of markers. The LD decay plot was plotted

Traits	Variance	Comb 1	Comb 2	Comb 3	Comb 4
Fe	Env	11.58 ***	10.18 **	24.601 **	1.08
	Env.Rep	0.24 **	0.24 *	0.097	0
	Gen	6.37 ***	7.4 ***	11.809 ***	19.03 ***
	Env.Gen	25.25 ***	27.37 ***	26.147 ***	25.85 ***
	Mean (ppm)	31.58	32.68	31.79	34.52
	Range (ppm)	24.26–40.5	24.01–43.03	22.09–50.7	23.37–52.55
	CV (%)	9.29	9.17	7.96	8.13
Zn	Env	11.74 ***	7.059 **	20.136 *	20.136
	Env.Rep	0.194 *	0.007	0.281 *	0.281
	Gen	8.669 ***	8.668 ***	12.484 ***	12.484 ***
	Env.Gen	14.069 ***	15.832 ***	10.954 ***	10.954 ***
	Mean (ppm)	29.39	30.72	28.58	32.27
	Range (ppm)	21.03–37.07	22.47–40.3	17.92–39.13	21.45–43.16
	CV (%)	9.56	9.07	8.38	7.32

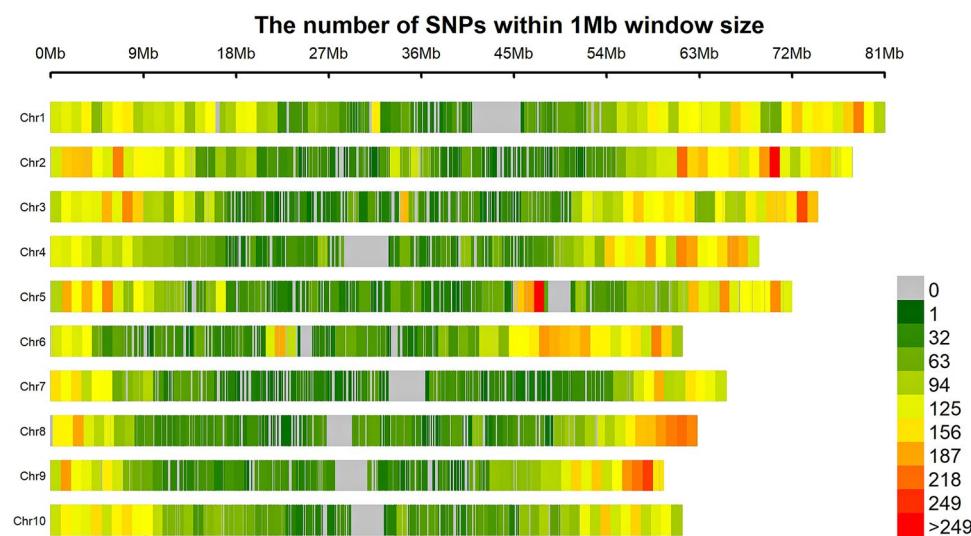
**Table 2.** ANOVA for variance components for both Fe and Zn traits for four combinations. \*, \*\* and \*\*\* denotes significant at the 5%, 1% and 0.1% probability level, respectively; Fe—Iron, Zn—Zinc, Comb—Combination of environments (as depicted in Table 1); Env—Environment; Env.Rep—Environment × Replications; Gen—Genotype; Env.Gen—Environment × Genotypes; ppm—parts per million (mg/kg); CV—coefficient of variation.



**Figure 1.** Relationship (Pearson correlation coefficient,  $r$ ) between grain iron and zinc across four different combinations of environments with a confidence interval (CI) of 95%. Comb\_1, \_2, \_3, \_4, respectively, environments combination 1, 2, 3, and 4.



**Figure 2.** Graphical representation of grain iron and zinc frequency distributions for all four combinations. Comb\_1, \_2, \_3, \_4, respectively, environments combination 1, 2, 3, and 4. Red color bars indicate the distribution of grain iron and blue color bars indicate the grain zinc distribution. The distribution of population is normal in all four combinations as described in Shapiro-Wilk test for normality in frequentist statistics. The Ryan-Joiner (RJ) statistic also confirms that phenotypic data follow a normal distribution.



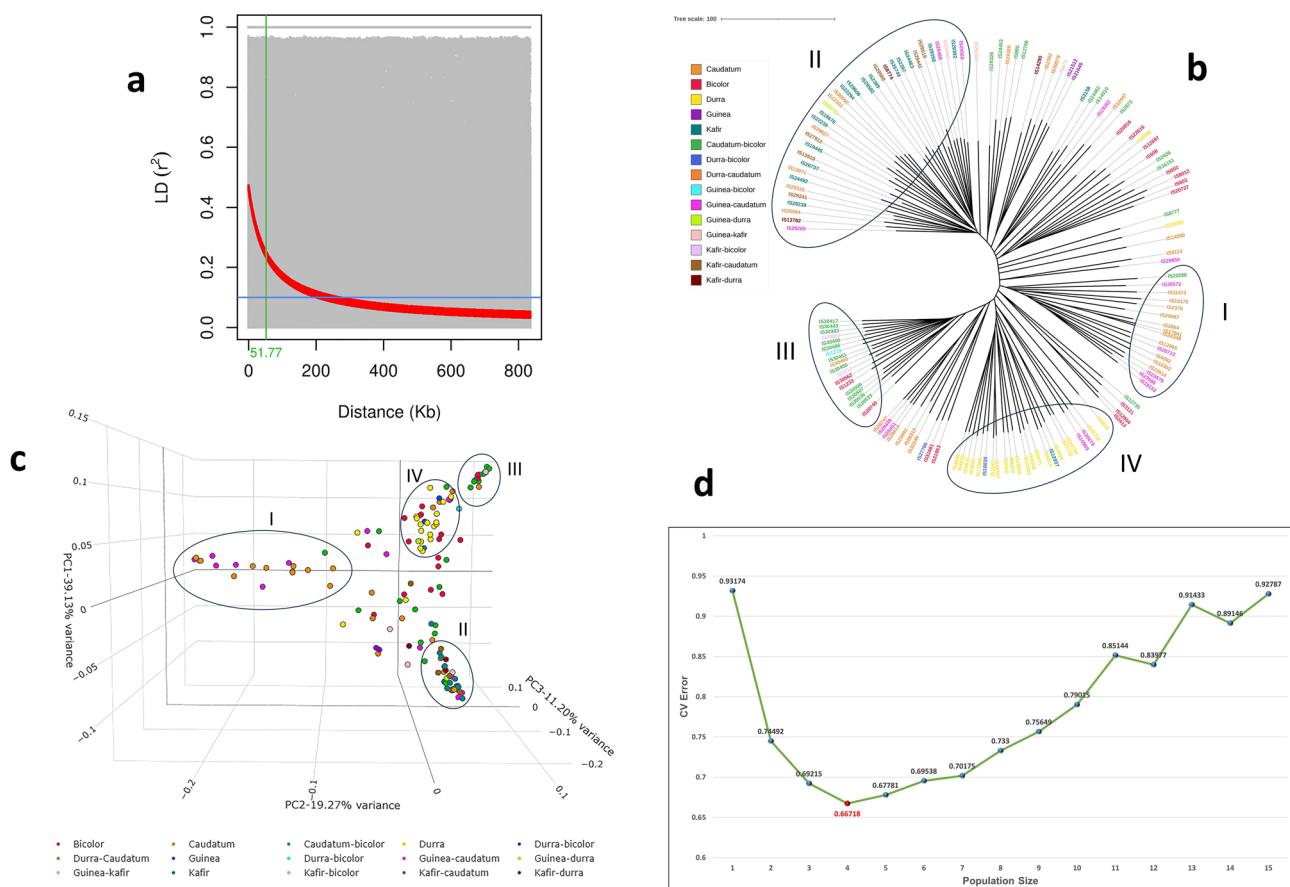
**Figure 3.** Distribution of SNP markers on ten chromosomes of the sorghum. Telomeric regions of the chromosomes are highly dense relative to the centromeric regions.

as LD ( $r^2$ ) against the distance in base pairs (bp). The overall LD decay across the genome estimated at 51.77 kb (Fig. 4a). The unweighted unrooted neighbor-joining tree (Fig. 4b) used to depict the phylogenetic diversity showed that the genotypes clustered based on the races, but also revealed a significant level of admixtures (Fig. 4b). A similar clustering was confirmed using the PCA with the first 3 PCs explaining 39.13%, 19.27% and 11.20% genetic variation (GVE), respectively (Fig. 4c), which gives a total of 70% GVE. The optimum population size estimation using the ADMIXTURE model also gave K=4 as the point with the lowest cross-validation error (CV error) (Fig. 4d).

### Marker-trait associations

A genome-wide association analysis was undertaken to use the whole-genome high-quality 55,068 marker information, high-repeatability phenotypic data, and four single-locus (SUPER) and multi-locus (BLINK, MLM, FarmCPU) algorithms, and evaluate the linkage disequilibrium (MTAs) that existed between genetic variations and nutritional traits viz., grain Fe and Zn concentration in the panel of 140 sorghum genotypes. For both Fe and Zn traits, five MTAs were detected, each, by BLINK and MLM GWAS models, whereas FarmCPU and SUPER, each, detected two. A threshold of MAF ( $\geq 0.05$ ) was set to account for the limitations of GWAS<sup>66</sup> and to correct for spurious MTAs. Table 3 displays a comprehensive compilation of significant markers that showed significant effects. The Manhattan plots (Fig. 5) showcased the GWAS output, portraying where these highly promising markers assert their statistical significance amidst the vast complexity of the genome. Additionally, these markers were carefully selected based on their elevated values in the Quantile-Quantile (QQ) plots (Fig. 5), indicating their non-random association with the traits of interest.

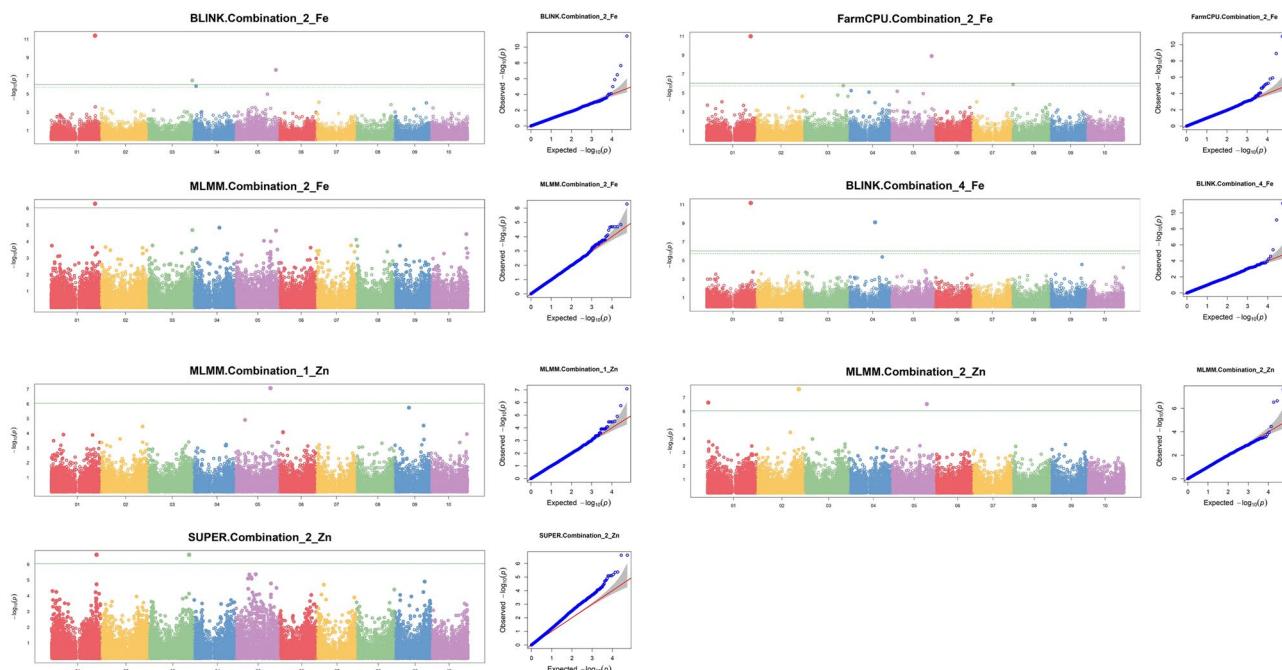
We identified 5 consistent MTAs for Zn and 4 for Fe that were revealed by at least one model in at least one trial combination (Table 3; Fig. 5). Three of the 5 MTAs for Zn were detected using the statistically stringent MLM model<sup>54,67</sup> while the remaining 2 were detected using the SUPER model. All Zn MTAs were detected for trial combination 2, except one (S05\_58213541), that was detected for trial combinations 1 and 2. The highest PVE reported for Zn was 31.8% (Table 3). The strongest MTA (S01\_72265728) reported for Fe was detected by 3



**Figure 4.** Informativeness of the markers and characterization of the structure of the genotypes used for GWAS. a. LD decay distance estimated at 51.77 Kb. b. A dendrogram showing the clustering of the genotypes used. Four clear clusters are observed with the rest appearing as admixtures. Cluster I—Caudatum and associated hybrids; Cluster 2—Kafir and associated hybrids; Cluster III—Caudatum-bicolor hybrids; Cluster IV—Durra genotypes. c. PCA showing informativeness of the markers and further confirming the clustering observed in the dendrogram. d. Optimum population size estimation confirms K=4 as the point with the lowest CV error.

Traits	SNP	P values				PVE (%)	Trial combinations
		MLMM	BLINK	SUPER	FarmCPU		
Fe	S01_72265728	5.19E-07 (2)*	3.88 E-12 (2); 6.47 E-12 (4)	N/A	9.94E-12 (2)	12.0–34.7	2,4
	S03_73164578	N/A	3.20E-07 (2)	N/A	N/A	8.8	2
	S04_43148417	N/A	7.67E-10 (4)	N/A	N/A	8.8	4
	S05_67287071	N/A	2.22E-08 (2)	N/A	1.22E-09 (2)	7.5; 14.7	2
Zn	S01_1495697	2.29E-07 (2)	N/A	N/A	N/A	13.8	2
	S01_73777110	N/A	N/A	2.44E-07 (2)	N/A	12.2	2
	S02_69057239	2.40E-08 (2)	N/A	N/A	N/A	16.5	2
	S03_67027187	N/A	N/A	2.46E-07 (2)	N/A	12.2	2
	S05_58213541	8.40E-08 (1); 2.96E-07 (2)	N/A	N/A	N/A	31.8 (1); 11.4 (2)	1,2

**Table 3.** The marker-trait associations (MTAs) or SNPs detected for grain iron (Fe) and grain zinc (Zn) using multiple model algorithms with MAF  $\geq 0.05$ . \*Numbers within round brackets refer to respective trial combinations in the rightmost (8th) column.



**Figure 5.** Manhattan plots along with their respective QQ-plots showing the association of sorghum accessions for grain Fe and Zn content. Manhattan (left) and respective QQ-plots (right) (from top to bottom) depicted for grain Fe using BLINK, FarmCPU and MLMM in combination 2; and combination 4 using BLINK; and for grain Zn combination 1 using MLMM; and combination 2 using MLMM and SUPER. Associations were detected using 55,068 high-quality SNPs. The green horizontal line in the Manhattan plot shows the Bonferroni threshold at a 5% level:  $-\log_{10}(p.\text{value}) < 6.04$ , above which solid dots indicated significant MTAs.

models including the most stringent, MLMM (Table 3). The same MTA was also detected across trial combinations 2 and 4, and was just 1.5 Mb downstream to a Zn MTA locus detected by marker S01\_73777110 (Table 3). The highest PVE for this Fe MTA was 34.7%. The other MTAs for Fe were detected by BLINK (S03\_73164578, S04\_43148417, S05\_67287071) and FarmCPU (S05\_67287071), all of which had the highest PVE at 14.7%. Supplementary Table S3 depicts further information about significant MTAs along with MAF, effects, genes, and the functional annotation.

### Candidate genes associated with grain Fe and Zn

Candidate genes for each of the MTAs were identified within the window of LD block (51.77 kb upstream–downstream) as defined by the LD decay estimation (Supplementary Table S4; Fig. 4a). Moreover, Table 4 provides a summary of all significant markers that were identified to be localized within genes, and the corresponding annotations. A total of 5 genes were identified for both traits studied (Table 4). Two significant and genic SNPs associated with Fe were linked with Sobic.001G445900 and Sobic.005G188300 genes which play a significant

Traits	SNP	Gene ID	Gene position	Sequence description	GeneOntologyID	GeneOntology	GO Biological process	GO Cellular component	GO Molecular function
Fe	S01_72265728	Sobic.001G445900. v3.2	72260202–72266394	similar to Cytochrome P450 heme binding iron ion binding oxidoreductase activity	GO:0005783 GO:0080132 GO:0020037 GO:0005506 GO:0004497 GO:0016491 GO:0016705 GO:0010012 GO:0016132 GO:0010268 GO:0009867 GO:0010358 GO:0009741 GO:0016125 GO:0009826	Endoplasmic reticulum fatty acid alpha-hydroxylase activity, heme binding, iron ion binding, monooxygenase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen steroid, 22-alpha hydroxy-lase activity, brassinosteroid biosynthetic process, brassinosteroid homeostasis, jasmonic acid mediated signaling pathway, leaf shaping response to brassinosteroid sterol, metabolic process unidimensional cell growth	Dhurrin biosynthetic process	Endoplasmic reticulum membrane, cytosol	Heme binding, iron ion binding, monooxygenase activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen
	S05_67287071	Sobic.005G188300. v3.2	67285702–67291437	Similar to Expressed Protein	N/A	N/A	N/A	N/A	N/A
Zn	S01_1495697	Sobic.001G017500. v3.2	1495068–1497208	Similar to Putative uncharacterized protein NAD(P)-binding domain Glucose/ribitol dehydrogenase	GO:0000166 GO:0008152 GO:0016114 GO:0016491 GO:0055114	NADPH-cytochrome P450 reductase, Inosine triphosphate triphosphate catabolic process, nucleotide metabolic process, NADH metabolic process, NADP metabolic process, NADP metabolic process	Deoxyribonucleoside triphosphate catabolic process, nucleoside triphosphate catabolic process, nucleotide metabolic process, NADH metabolic process, NADP metabolic process	Cytosol, endoplasmic reticulum membrane, cytoplasm, cytoskeleton, amyloplast	Flavin adenine dinucleotide binding, FMN binding, NADP binding, NADPH-hemoprotein reductase activity, metal ion binding, ATP binding, hydrolase activity
	S01_73777110	Sobic.001G463800. v3.2	73772286–73782185	Similar to Expressed Protein of unknown function DUF566	GO:0005737 GO:0005880 GO:0008017 GO:0051225	Malate dehydrogenase [NADP] 1, chloroplastic, Cyanhydrin beta-glucosyltransferase, Phosphoenolpyruvate carboxylase 3	Malate metabolic process, dhurrin biosynthetic process, carbon fixation, photosynthesis, tricarboxylic acid cycle	Cytoplasm, endoplasmic reticulum membrane	Malate dehydrogenase (NADP+) activity, cyanhydrin beta-glucosyltransferase activity, magnesium ion binding, iron ion binding
	S03_67027187	Sobic.003G350800. v3.2	67025951–67028277	Expressed Protein HSP20-like chaperone	GO:0042538 GO:0009941	Peptidase S8/S53 domain-containing protein,	Hyperosmotic salinity response, proteolysis, chloroplast organization, heat acclimation	Golgi apparatus, membrane, cytoplasm	Serine-type endopeptidase activity, alpha-mannosidase activity, carbohydrate binding, metal ion binding, RNA cap binding

**Table 4.** A summary of SNP markers located within a gene and the corresponding annotation.

role in Cytochrome P450, heme binding, iron ion binding, oxidoreductase activity, and monooxygenase activity. The genes are putatively involved in other processes including acting on paired donors with incorporation or reduction of molecular oxygen in dhurrin biosynthetic process, jasmonic acid mediated signaling pathway, leaf shaping response to brassinosteroid sterol, and metabolic process unidimensional cell growth. Similarly, three significant SNPs associated with Zn were located in Sobic.001G017500, Sobic.001G463800 and Sobic.003G350800 genes which are putatively associated several properties including NAD(P)-binding domain, Glucose/ribitol dehydrogenase, NADPH-cytochrome P450 reductase, peptidase S8/S53 domain-containing protein and malate dehydrogenase acting in deoxyribonucleoside triphosphate catabolic process, nucleoside triphosphate catabolic

process, nucleotide metabolic process, NADH metabolic process, carbon fixation, photosynthesis, magnesium ion binding and metal ion binding.

### Haplotype block analysis

In the analysis of trait-associated markers, an adjacent set of five markers was utilized to assess the haplotype block for each marker. This approach resulted in an average block size of 109 kb for iron (Fe) trait markers and 66 kb for zinc (Zn) trait markers. The analysis identified a total of eleven accessions exhibiting promising characteristics for Fe traits, and fifteen accessions demonstrating favorable attributes for Zn traits. Notably, five accessions viz., IS22294, IS29239, IS22239, IS29565, and IS23514 were found to possess haplotypes blocks of markers associated with both Fe and Zn traits (Fig. 6). Additionally, based on the Zn haploblock analysis, three accessions (IS22239, IS21512, and IS19676) demonstrated unique pattern of haplotypes indicating that these are related and have the strong variation in Block 1.

### Expression analysis

The tissue-specific expression of the identified genes was analyzed in 36 sorghum tissues at three different growth stages of the plant (Supplementary Table S5, Fig. 7). Based on the FPKM values, an expression heatmap was generated, plotting the expression data of five candidate genes across 36 sorghum tissues (Supplementary Table S5, Fig. 7). Out of five genes, Sobic.003G350800 which is a peptidase S8/S53 domain-containing protein, showed higher level of expression in several tissues such as leaf, root, flower, panicle, and stem. Conversely, Sobic.001G017500 is not expressed in any of the tissues. Sobic.005G188300 and Sobic.001G463800 genes are expressed moderately at grain maturity and anthesis stages in leaf, root, panicle, and seed tissues. Sobic.001G445900 gene showed high expression at juvenile and grain maturity stage.

### Discussion

Iron (Fe) and zinc (Zn) are essential micronutrients for both human and animal nutrition, and their importance in sorghum lies in their role in promoting overall health and preventing nutrient deficiencies, also known as hidden hunger. In addition to preventing micronutrient malnutrition, these micronutrients are essential for cognitive development, immune system function, agricultural productivity, and socio-economic development. Hence,

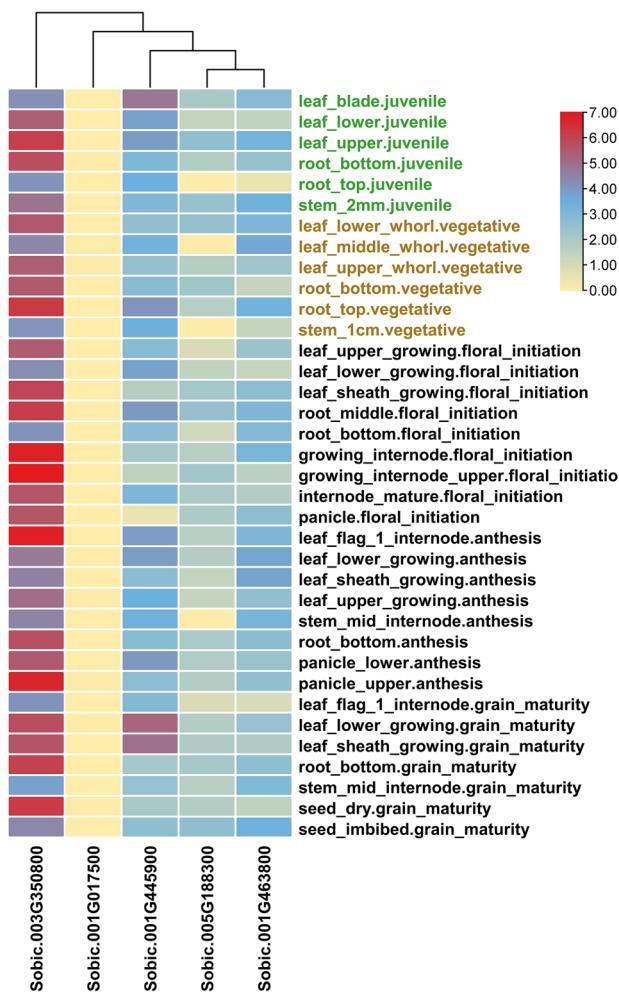
#### Fe trait associated haploblocks

	Block 1	Block 2	Block 3	Block 4
	S01_72265728	S03_73164578	S04_43148417	S05_67287071
	↓	↓	↓	↓
IS22294	G C G A C T T C G A T A N A G C A A A N C T G W C C C A   A G G A N T T G C C G   G G A C C			
IS29239	G C G A C T T C G A T A G A G C A A A G C T G W C C C W A G G A T T A C C A C G G A C C			
IS19389	G C G A C T T C G A T A G A G C A A A G N T G T C C C A   A G G A T T A C C A C G G A C C			
IS12447	N C G A C T T C G A T A G A G C A A A G C T G T C C C N A G G W T T A C C A C G G A C C			
IS22239	G C G A C T T C G A T A G A G C A A N R C T G T C C C W A G G A T T A C C A C N G A C N			
IS29565	G C G A C T T C G A T A G A G C A A A G C T G W C C C A   A G G A T T N N C A C G G A C C			
IS23514	G C G A C T T C G A T A N A A G C A A A G C T N A C C C T A G G T T T A C C A C G G A C C			
IS24492	G C G A C T T C G A T A G A G C A A A G C T G A C C C A   A G G T T T N N C A C G G A C C			
IS24463	G C G A C T T C G A T A G A G C A A N G C T G W C C C A   A G G W T T A C C A C G G A C C			
IS29950	G C G A C T T C G A T A T A G A G C A A A A A T G W C C C A   A G G A T T A C C A S G G A C C			
IS20298	G C G A C T T C G A N A G A G C A A A G C T N W C C C A   A G G A T T A A C A C G G A C G			
BlockPattern	G C G A C T T C G A T A T A G C A A A G C T G T C C C A A G G A T T T G C C A C G G A C C			

#### Zn trait associated haploblocks

	Block 1	Block 2	Block 3	Block 4	Block 5
	S01_1495697	S01_73777110	S02_69057239	S03_67027187	S05_58213541
	↓	↓	↓	↓	↓
IS22294	G T G T G R T A T T G T C A C T T G G A G T A C A T T G G   G A G G C G T C C T C   G A G C C A T A C C T G Y C				
IS29565	T N C G C R T A T T G T C A C T T G A A C A C A T T G G   G A G G C G T C N T C   G A G C Y G T A G C T G N N N C				
IS29269	G Y C G C N T A T T G T C A C T T G A A C A C A T T G G   G A G G G N G T C C T C   G A G C Y G T A G C T G N N N C				
IS22239	G N G T G A G C C G T T C A C T T G A A C A C A T T G G   G A G G G N G T C C T C   G A G C Y G T A G C T G T C T C				
IS30092	T C C G G C A T A T T G T C A C T T G A A C A C A T T G G   G A G G G N G T C C T C   G A G C C A T A C C T G C C T C				
IS23514	T C C G G C A T A T T G G C C A A C C G G T T G T G A G G   G A G T N G T C C T C   G A G C C G T A G T G T T G C C T C				
IS29239	G C C G G C A T A T T G T C A C T T G A A C A C A T T G G   G A G G G N G T C C T C   G A G C C G T A G T G C T K T C Y C				
IS33090	G C C G G C A T A T T G T C A C A C C G A A T A T T G G   G A G G T C G T C T G A A A G A C C G N A G C T T G T C Y C				
IS30079	G T G T G R T A T T G T C A C T T G A A T A C A T T A A A G C G C G T C C T C   G A G C C A T A C C T T G C C T C				
IS29606	G T G T G R T A T T G T C A C T T G A A N A C A T T G G   G A G G G C G T C C T C   G A G C C A T A C C T T G C C T C				
IS29392	T N C G C A T A T T G T C A C T T G A A C A C A T T G G   G A G G G N G T C C N C   G A G C C G T N G C T G T C T T N				
IS31681	T C C G G C G T A T T G T C A A T C G A A T A C A T T G G   G A G G G Y C C G T G A A G C A C G T G G G K G T Y T C				
IS32787	G C C G G C A T A T T G G N N G A C C G A A T A C A T T G G   G A G G G C G T C T G A A N G A C G N G G C G G T T T M				
IS21512	N T G T G A G C C G T T C A C T T G A A C A C A T T G G   G A G G G N G T C C K C   G A G C C G T N G C T G T C C T C				
IS19676	G T G T G R G C C G T T C A C T T G A A C A C A T T G G   G A G G G C G T C C T C   G A G C C A T A C C T G C C T C				
BlockPattern	G C C G G C A T A T T G T C A A C C G A A T A C A T T G G G A G G G C G T C C T C   G A G C C G T A G C T G T C T C				

**Figure 6.** Haplotype blocks for the significant MTAs for traits studied. Five adjacent markers on both sides of the significant SNP were used to find the significant haplotype block. The ‘BlockPattern’ at the end of the haplotype blocks describes the pattern of the respective hapblock.



**Figure 7.** Tissue-specific expression of candidate genes identified for Fe and Zn contents. The expression of five candidate genes in 36 tissues from three different growth stages (juvenile, vegetative, and reproductive) are plotted in the heatmap. Juvenile, vegetative, and reproductive stages are depicted in green, brown, and black font color, respectively.

biofortified sorghum provides a readily available and affordable source of essential micronutrients, preventing and alleviating micronutrient deficiencies and preserving traditional food culture. It preserves traditional food preferences, promotes economic development, and contributes to climate resilience as sorghum is a C4 crop with high level of resource use efficiency. Few biofortified sorghum varieties were released under the world's drylands and this gap motivated our research work; we aimed to identify the genomic regions associated with grain Fe and Zn for downstream use in sorghum biofortification breeding. Several GWAS studies has identified the genes governing complex traits, revolutionizing agricultural improvement. Research on pulses and grains revealed genes related to the accumulation of micronutrients, including the pathway for the carotenoid production in maize<sup>68</sup>. Additionally, wheat displayed 92 SNP trait correlations linked to 10-grain mineral areas. Twenty genes were functionally annotated to demonstrate their significance in grain mineral accumulation; most of these genes were found in the D-genome, indicating control over wheat grain mineral diversity<sup>69</sup>.

The genetic architecture of seed molybdenum and selenium in wild and cultivated chickpea was investigated using GWAS<sup>70</sup>. After 180 entries were surveyed, 16 SNPs were linked to these characteristics. Similarly, 22 quantitative trait nucleotides (QTNs) for grain nitrogen were found by phenotyping 174 accessions of Croatian common bean land races for the seed concentrations of eight micronutrients<sup>71</sup>. In our study, using sorghum entries from the ICRISAT's minicore provided a guarantee that the uncovered major QTLs and proxy SNPs would benefit wider research communities. To the best of our knowledge, this work is the first to report on the marker traits associations for grain Fe and Zn in sorghum using SNPs and minicore lines.

In this study, the analysis of variance (ANOVA) was carried out for both the characters viz. grain Fe and grain Zn over four combinations of environments. The result indicates highly significant differences among the genotypes for both traits, which revealed the existence of sufficient variation for effective GWAS, statistical inferences, and selection of superior plant ideotypes. Despite the study's small population size, the identified MTAs were significant and agreed with previous results, indicating that the population size did not pose severe data

quality challenges<sup>33,72,73</sup>. This study demonstrated that GWAS is a powerful tool for identifying potential genetic factors that contribute to important traits in sorghum genotypes, even with a small sample size.

The repeatability for different combinations was varied, the highest was measured for Fe and Zn (56% and 74%, respectively) in combination 4, whereas low values were registered in combination 2 i.e., 41% and 57% for Fe and Zn, respectively. The genetic basis of traits like grain Fe and Zn content can be complex, involving multiple genes with small effects<sup>24</sup>, and this can explain the observed moderate repeatability. The observed medium to high broad-sense heritability/repeatability and wide variation within population for the evaluated micronutrients are the precondition for a successful GWAS<sup>4–76</sup>. The association between grain Fe and Zn showed significant and high positive values and the trend was similar across environments, implying that selection for either mineral can be used as proxy for the other. Such a favorable relationship between these micronutrients were reported earlier in sorghum<sup>77–80</sup> and other cereals, such as pearl millet<sup>81–86</sup>, maize<sup>87,88</sup>, rice<sup>89,90</sup>, and wheat<sup>91–94</sup>.

Research on sorghum grain Fe and Zn so far could only identify a few small effect QTLs and some putative candidate genes<sup>24</sup>, without meaningful use in sorghum genetic biofortification. In this study we used multiple popular GWAS models i.e., MLM, BLINK, FarmCPU and SUPER<sup>33,37</sup>; the latter is a single locus model, while the remaining three are multi-locus. The use of multiple models allowed to detect more reliable QTLs i.e., those that were co-detected by more than one model<sup>95</sup>. On the other hand, although multi-locus GWAS models showed advantage over single-locus GWAS methods<sup>96</sup>, a combination of single-locus methods and multi-locus methods was used in this work as recommended by<sup>97–100</sup> to improve the detection power and robustness of GWAS. The FarmCPU method offers superior statistical power by dividing Multiple Loci Linear Mixed Models into fixed effects model (FEM) and random effects model (REM), removing confounding, and controlling false positives; the SUPER model addresses computing issues with MLM, while BLINK improves statistical power with using LD information<sup>33,54</sup>. Structure analyses showed that the minicore population was genetically structured with an estimated four subpopulations, and corrective measures were implemented in the GWAS to account for population structure and cryptic relationships to avoid false positive associations<sup>33,101,102</sup>.

With the use of GBS SNP data, we identified nine highly significant MTAs for grain Fe and Zn with P values ranging from  $3.88 \times 10^{-12}$  to  $5.19 \times 10^{-7}$ , explaining 7.5 to 34.7% of phenotypic variation (PVE). Four MTAs for Fe were identified on chromosomes 01, 03, 04 and 05 while five MTAs for Zn were identified on chromosomes 01, 02, 03 and 05. Two major SNPs were identified for Fe and Zn: for iron, the SNP S01\_72265728 was identified in the cytochrome P450 gene, while the SNP S05\_58213541 associated with Zn is intergenic and near Sobic.005G134800 (2.8 kb) which codes for a zinc binding ribosomal protein. S01\_72265728 was associated with positive effect on Fe accumulation in the kernel, while S05\_58213541 was associated with negative effect on Zn accumulation. Many other differentially expressed genes are involved in the uptake of minerals. In the case of the gene for Fe, the Cytochrome P450 gene is involved in mineral uptake for both Fe and Zn. Cytochrome P450 is a pigment with heme-protein properties and participates in several catalytic processes that involve specific heme group<sup>103</sup>. Cytochrome P450s catalyze a wide range of chemical reactions and have different enzymatic mechanisms and complex substrate specificities. The enzyme structures of different P450 possess a heme-binding mode with an unusually long heme-binding loop and a unique I-helix which may be involved in Fe uptake from the soil<sup>104</sup>. A study on barley collection identified single-nucleotide polymorphisms for Fe and Zn biofortification in which cytochrome P450 superfamily protein was found to be involved in element transport, iron, and zinc binding<sup>43</sup>. Plant cytochrome P450 (P450) participates in a wide range of biosynthetic reactions and targets a variety of biological molecules<sup>105</sup>. However, there is no specific information on how cytochrome P450 gene is involved in Fe and Zn uptake. Satyavathi et al.<sup>106</sup> identified a cytochrome P450 superfamily protein involved in element transport, iron, and zinc binding, and found out that it was present in the genotypes with high Fe and Zn contents. A study on rice biofortification with zinc and selenium found that the expression pattern of a Cytochrome P450 (CYP) gene followed the mineral accumulation in flag leaves<sup>107</sup>. Such mixed reports suggest that more research is needed to fully understand the function of cytochrome P450 gene in mineral uptake. In pearl millet, it was reported that the cytochrome P450 proteins are up-regulated during panicle initiation<sup>105</sup>. The process of iron uptake by plants is an extremely energy-intensive mechanism<sup>108</sup>. A plant's ability to extract iron from the complex or chelating molecule by reducing Fe+++ to Fe++ is essential for iron absorption<sup>108</sup>. Uptake of Fe from the soil is dependent on other cations in the soil solution such as manganese (Mn) and calcium (Ca). Accumulation of Fe in grain is a very complex process and many factors such as polyphenol content and other stresses play a significant role in Fe concentration in the grain. In another study found that drought stress alters iron accumulation in sorghum seeds, and photosynthesis impaired by drought stress might trigger a disturbance in iron homeostasis<sup>109</sup>. The same authors hypothesized that increased vacuolar transporters and ferritin might be involved in the regulation of iron accumulation in sorghum seeds under drought stress.

Zinc-binding proteins are involved in abiotic stress tolerance and play a significant role in root hair growth under stress conditions. It was reported that the metal binding proteins facilitate the absorption of Zn and metal ions by procuring the binding sites<sup>110</sup> but in our study, the identified SNP (S05\_58213541) showed a negative effect which could mean that there is an association between that SNP and a particular trait, but the presence of a specific allele at that SNP locus is associated with a decrease or reduction in the trait of interest, Zn concentration in this case. This can indicate that either the plant absorbed the Zn ions from the soil and sequestered it but unable to translocate it to the grains, or the plant with that particular SNP was unable to uptake enough Zn from the soil. Whether this is a direct effect (modified gene product) or a non-allelic interaction, it is to be investigated. This gene can also be a candidate for genome editing in order to improve Zn concentration in the kernel. If the associated gene is knocked out, there may be chances of significantly increasing the concentration of Zn in grains. Most putative genes identified in this study are zinc-binding or zinc ion-binding proteins. The gene ontology CCHH term indicates the presence of transcripts involved in metal ion binding activity, indicating their role in uptake and transport of Fe and Zn. A zinc finger is a small protein structural motif characterized by the coordination of

one or more zinc ions ( $Zn^{2+}$ ) to stabilize the fold. Zinc finger proteins are transcription factors with the finger domain, which plays a significant role in gene regulation and are required for transcriptional activation.

Based on the results, among the four combinations of locations, combination 2 consisting of three locations identified the most MTAs. The identified candidate genomic regions/candidate genes are likely to have an important role in achieving high Fe and Zn content in sorghum grains. The identified SNPs can therefore be validated and used in developing nutritionally improved, Fe and Zn rich sorghum cultivars, which would help address micronutrient deficiencies and increase food security. The correlation between Fe and Zn concentration in the sorghum grain was high ( $r > 0.80$ ) but we could not come across pleiotropic QTLs (SNPs) in this study. The lack of common SNPs in this study may indicate that the effects of the genetic factors that influence Fe and Zn concentrations may vary depending on the environmental conditions. Indeed, a correlation of 0.80 implies that 36% of the variance in sorghum grain Fe or Zn concentration cannot be explained by either of the two metals.

Haplotype-based breeding has recently come to prominence as an effective approach to developing crop varieties that meet particular requirements<sup>111</sup>. To be used in breeding programs, this breeding strategy needs to first determine superior haplotypes. Haplotypes are distinct sets of alleles found on a single chromosome that are inherited together with a limited probability of contemporary recombination<sup>112</sup>. Breeders can increase the accuracy of genomic predictions and breeding strategies by more effectively defining haplotypes using linkage disequilibrium-based techniques and haplotype diversity<sup>113,114</sup>.

Recently, researchers explained the genetic diversity and evolutionary background of sorghum accessions by identifying several haplotypes of genes such as *Dry*<sup>115</sup> and *Sh1*<sup>116</sup> in domesticated and wild sorghum lines. Furthermore, Wu et al.<sup>117</sup> conducted a population genetic study that demonstrated the *Sh1* and *SbTB1* regions were subject to strong selection during the domestication of sorghum. Moreover, it has resulted in the potential role of *SbTB1* haplotype in controlling the number of lateral branches in sorghum in domesticated and wild accessions. To the best of our knowledge, there is no specific study that directly addresses haplotype analysis in grain Fe and Zn in sorghum, and this is the first study to report haplotypes for grain Fe and Zn. In the present investigation, the haplotype analysis used sets of five markers to evaluate haplotype blocks, resulting in average block sizes of 109 kb for iron (Fe) trait markers and 66 kb for zinc (Zn) trait markers. It identified eleven accessions with promising Fe traits and fifteen with favorable Zn traits. The identified haploblocks were further used to identify sorghum accessions that inherited iron and/or zinc QTLs stably and possibly with rare crossing-over events. Notably, five accessions possessed haplotype blocks associated with both Fe and Zn traits. Furthermore, three accessions showed unique haplotype patterns related to strong variation in Block 1 based on Zn haploblock analysis. Thus, it will be beneficial to: (1) target the identified haploblock-containing plants as potential QTL donors in sorghum crossing blocks, and (2) tailor plant architecture and enhance grain nutrient content by incorporating these haplotypes through haplotype-based breeding in sorghum breeding programs.

In this work, the baseline expression profiles of the identified genes were shown using heat map. The gene Sobic.003G350800 encoding for a protein which contains the peptidase S8/S53 domain, was found to be highly expressed in a variety of tissues, including the leaf, root, flower, panicle, and stem. Nearly all plant species include peptidase S8/S53 domain subtilases, which are subtilisin-like proteases that regulate a variety of biotic and abiotic stressors, and are expressed in a variety of tissues. *Sorghum bicolor* contains 57 genes belonging to the S8 family<sup>118</sup>. Subtilase family members exhibit remarkable functional versatility, from protein expression to stress responses, and modulating plant growth and development from seed development to senescence<sup>119</sup>. Genes, Sobic.005G188300 and Sobic.001G463800 are expressed moderately at grain maturity and anthesis stages in leaf, root, panicle, and seed tissues. These genes mainly encode for malate dehydrogenase (MDH: EC 1.1.1.82) in C4 plants. The functional annotation of Sobic.001G463800 showed its role in iron ion binding activity, which is probably one of the factors determining the success of biofortification, nutrient bioavailability at different plant growth stages, from soil to plant tissues. In nature, the synthesis of malate is catalyzed by the MDH enzyme through the reversible reduction of oxaloacetate to malate. As such, MDH is therefore indirectly but strongly associated with iron and other nutrients uptake in plants. Indeed, Malate is a key product of plant metabolism with diverse functional roles in plants<sup>120,121</sup>, including, but not limited to: respiration and energy generation, photosynthesis (both C3 and C4), fatty acid oxidation, lignin biosynthesis, pulvinal and stomatal function, nitrogen (N2) fixation and amino acid biosynthesis, ion balance, uptake of phosphorus (P) and iron (Fe), and aluminum (Al) tolerance<sup>121–123</sup>. Cultivar differences, such as low nutrient mobility and remobilization efficiency, also affect the effectiveness of biofortification, particularly in relation to leaves and edible parts<sup>124</sup>. The identified candidate genes showing high micronutrients expression in leaves, stem, and grains at different growth stages (juvenile, vegetative, and reproductive stages) can be targeted to increase the nutritional value of both grain and stover, providing health-promoting food, feed, and forage. Overall, this study provides invaluable information on the genetic basis of grain Fe and Zn contents in sorghum and identifies candidate genes and genomic regions that can be used in breeding programs to improve these micronutrients. However, more research works are needed to further characterize these genomic regions and candidate genes and to validate their role in grain Fe and Zn concentration in sorghum grain before they are implemented in sorghum breeding programs.

## Data availability

The datasets created during and/or analyzed during the current investigation are not publicly available, although the corresponding author can provide them upon justifiable request.

Received: 26 December 2023; Accepted: 27 May 2024

Published online: 03 June 2024

## References

- FAO. World Food and Agriculture—Statistical Yearbook 2021. Rome (2021). <https://doi.org/10.4060/cb4477en>.
- WHO. Malnutrition. <https://www.who.int/news-room/fact-sheets/detail/malnutrition> (2021a). Accessed 4 Sept 2023.
- WHO. Children: improving survival and well-being. <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality#:~:text=Malnourished%20children%2C%20particularly%20those%20with,under%2D5%20years%20of%20age> (2020). Accessed 21 Aug 2023.
- WHO. *World Health Statistics 2022: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO. ISBN 978-92-4-005114-0 (2022).
- WHO. *World Health Statistics 2023: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO (2023a).
- Ingle, K. P. *et al.* Biofortified foxtail millet: towards a more nourishing future. *Plant Growth Regul.* **99**, 25–34. <https://doi.org/10.1007/s10725-022-00900-2> (2023).
- Abbaspour, N., Hurrell, R. & Kelishadi, R. Review on iron and its importance for human health. *J. Res. Med. Sci.* **19**(2), 164–174 (2014).
- Piskin, E., Cianciosi, D., Gulec, S., Tomas, M. & Capanoglu, E. Iron absorption: factors, limitations, and improvement methods. *ACS Omega* **7**(24), 20441–20456. <https://doi.org/10.1021/acsomega.2c01833> (2022).
- WHO. Anaemia. Available from: <https://www.who.int/news-room/fact-sheets/detail/anaemia>. Accessed 25th Sept 2023 (2023b).
- WHO. Anaemia in women and children. [https://www.who.int/data/gho/data/themes/topics/anaemia\\_in\\_women\\_and\\_children#:~:text=In%202019%2C%20global%20anaemia%20prevalence,39.1%25\)%20in%20pregnant%20women](https://www.who.int/data/gho/data/themes/topics/anaemia_in_women_and_children#:~:text=In%202019%2C%20global%20anaemia%20prevalence,39.1%25)%20in%20pregnant%20women) (2021b). Accessed 04th Sept. 2023.
- Saper, R. B. & Rash, R. Zinc: an essential micronutrient. *Am. Fam. Physician* **79**(9), 768–772 (2009).
- McClung, J. P. Iron, zinc, and physical performance. *Biol. Trace Elem. Res.* **188**, 135–139. <https://doi.org/10.1007/s12011-018-1479-7> (2019).
- Fallah, A., Mohammad-Hasani, A. & Colagar, A. H. Zinc is an essential element for male fertility: A review of Zn roles in men's health germination, sperm quality, and fertilization. *J. Reprod. Infertil.* **19**(2), 69–81 (2018).
- Sanna, A., Firini, D., Zavattari, P. & Valera, P. Zinc status and autoimmunity: A systematic review and meta-analysis. *Nutrients* **10**(1), 68. <https://doi.org/10.3390/nu10010068> (2018).
- Skalny, A. V., Aschner, M. & Tinkov, A. A. Zinc. *Adv. Food Nutr. Res.* **96**, 251–310. <https://doi.org/10.1016/bs.afnr.2021.01.003> (2021).
- Hambidge, K. M. & Krebs, N. F. Zinc deficiency: A special challenge. *J. Nutr.* **137**(4), 1101–1105. <https://doi.org/10.1093/jn/nzj137.4.1101> (2007).
- Hussain, A. *et al.* Mechanistic impact of zinc deficiency in human development. *Front. Nutr.* **9**, 717064. <https://doi.org/10.3389/fnut.2022.717064> (2022).
- Narváez-Caicedo, C., Moreano, G., Sandoval, B. A. & Jara-Palacios, M. Á. Zinc deficiency among lactating mothers from a peri-urban community of the Ecuadorian andean region: An initial approach to the need of zinc supplementation. *Nutrients* **10**(7), 869. <https://doi.org/10.3390/nu10070869> (2018).
- DeFries, R. *et al.* Climate resilience of dry season cereals in India. *Sci. Rep.* **13**, 9960. <https://doi.org/10.1038/s41598-023-37109-w> (2023).
- Aruna, C. & Madhusudhana, R. Breeding for yield and adaptation. In *Sorghum in the 21st Century: Food—Fodder—Feed—Fuel for a Rapidly Changing World* (eds Tonapi, V. A. *et al.*) (Springer, Singapore, 2020). [https://doi.org/10.1007/978-981-15-8249-3\\_15](https://doi.org/10.1007/978-981-15-8249-3_15).
- Khoddami, A. *et al.* Sorghum in foods: Functionality and potential in innovative products. *Crit. Rev. Food Sci. Nutr.* **63**(9), 1170–1186. <https://doi.org/10.1080/10408398.2021.1960793> (2023).
- Sorghum Market Outlook 2022—2026. (2022). Global Sorghum Trends in 2022. <https://www.reportlinker.com/clp/global/48286#~:text=Global%20Sorghum%20Trends%20in%202022&text=Global%20consumption%20of%20sorghum%20is,with%205.7%20million%20metric%20tons>
- Kumar, A. A. *et al.* Recent advances in sorghum biofortification research. In *Plant Breeding Reviews* (ed. by Jules, J.), Vol. 39. <https://doi.org/10.1002/9781119107743.ch3> (2015).
- Kotla, A. *et al.* Identification of QTLs and candidate genes for high grain Fe and Zn concentration in sorghum [*Sorghum bicolor* (L.) Moench]. *J. Cereal Sci.* <https://doi.org/10.1016/j.jcs.2019.102850> (2019).
- Ashok Kumar, A., Reddy, B. V. S., Ramaiah, B., Sahrawat, K. L. & Pfeiffer, W. H. Genetic variability and character association for grain iron and zinc contents in sorghum germplasm accessions and commercial cultivars. *Eur. J. Plant Sci. Biotech.* **6**(Special Issue 1), 66–70 (2012).
- Madhusudhana, R. *et al.* Genetic variability, G × E interaction and stability for iron and zinc content in sorghum grains in advanced breeding lines. *J. Cereal Sci.* <https://doi.org/10.1016/j.jcs.2023.103653> (2023).
- Hariprasanna, K., Agte, V. & Patil, J. V. Genotype × environment interactions for grain micronutrient contents in sorghum [*Sorghum bicolor* (L.) Moench]. *Indian J. Genet. Plant Breed.* **72**(4), 429–434. <https://doi.org/10.5958/j.0975-6906.74.1.006> (2012).
- Garg, M. *et al.* Biofortified crops generated by breeding, agronomy, and transgenic approaches are improving lives of millions of people around the world. *Front. Nutr.* **5**, 12. <https://doi.org/10.3389/fnut.2018.00012> (2018).
- Gaddameedi, A. *et al.* The location of iron and zinc in grain of conventional and biofortified lines of sorghum. *J. Cereal Sci.* <https://doi.org/10.1016/j.jcs.2022.103531> (2022).
- Girma, G. *et al.* A large-scale genome-wide association analyses of Ethiopian sorghum landrace collection reveal loci associated with important traits. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.00691> (2019).
- Enyew, M. *et al.* Genome-wide analyses using multi-locus models revealed marker-trait associations for major agronomic traits in *Sorghum bicolor*. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2022.999692> (2022).
- Wondimu, Z., Dong, H., Paterson, A. H., Worku, W. & Bantte, K. Genome-wide association study reveals genomic loci influencing agronomic traits in Ethiopian sorghum (*Sorghum bicolor* (L.) Moench) landraces. *Mol. Breed.* <https://doi.org/10.1007/s11032-023-01381-5> (2023).
- Habyarimana, E., Dall'Agata, M., De Franceschi, P. & Baloch, F. S. Genome-wide association mapping of total antioxidant capacity, phenols, tannins, and flavonoids in a panel of *Sorghum bicolor* and *S. bicolor* × *S. halepense* populations using multi-locus models. *PLoS ONE* **14**(12), e0225979. <https://doi.org/10.1371/journal.pone.0225979> (2019).
- Rhodes, D. H. *et al.* Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. *J. Agric. Food Chem.* **62**(45), 10916–10927. <https://doi.org/10.1021/jf503651t> (2014).
- Rhodes, D. H. *et al.* Genetic architecture of kernel composition in global sorghum germplasm. *BMC Genom.* <https://doi.org/10.1186/s12864-016-3403-x> (2017).
- Shakoor, N. *et al.* Integration of experiments across diverse environments identifies the genetic determinants of variation in sorghum bicolor seed element composition. *Plant Phys.* **170**(4), 1989–1998. <https://doi.org/10.1104/pp.15.01971> (2016).
- Habyarimana, E., De Franceschi, P., Ercisli, S., Baloch, F. S. & Dall'Agata, M. Genome-wide association study for biomass related traits in a panel of *Sorghum bicolor* and *S. bicolor* × *S. halepense* populations. *Front. Plant Sci.* **11**, 551305. <https://doi.org/10.3389/fpls.2020.551305> (2020).

38. Pujar, M. *et al.* Genome-wide association study uncovers genomic regions associated with grain iron, zinc and protein content in pearl millet. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-76230-y> (2020).
39. Anuradha, N. *et al.* Deciphering genomic regions for high grain iron and zinc content using association mapping in pearl millet. *Front. Plant Sci.* **8**, 412. <https://doi.org/10.3389/fpls.2017.00412> (2017).
40. Zhou, Z. *et al.* Identification of novel genomic regions and superior alleles associated with Zn accumulation in wheat using a genome-wide association analysis method. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21061928> (2020).
41. Krishnappa, G. *et al.* Identification of novel genomic regions for biofortification traits using an SNP marker-enriched linkage map in wheat (*Triticum aestivum* L.). *Front. Nutr.* <https://doi.org/10.3389/fnut.2021.669444> (2021).
42. Hindu, V. *et al.* Identification and validation of genomic regions influencing kernel zinc and iron in maize. *Theor. Appl. Genet.* **131**, 1443–1457. <https://doi.org/10.1007/s00122-018-3089-3> (2018).
43. Nyiraguhirwa, S. *et al.* A genome-wide association study identifying single-nucleotide polymorphisms for iron and zinc biofortification in a worldwide barley collection. *Plants* **11**(10), 1349. <https://doi.org/10.3390/plants11101349> (2022).
44. Talukdar, P. *et al.* Identification of genomic loci regulating grain iron content in aus rice under two irrigation management systems. *Food Energy Secur.* <https://doi.org/10.1002/fes3.329> (2022).
45. Upadhyaya, H. D., Vetriventhan, M. & Deshpande, S. Sorghum germplasm resources characterization and trait mapping. In *The Sorghum Genome. Compendium of Plant Genomes* (eds Rakshit, S. & Wang, Y. H.) (Springer, Cham, 2016).
46. Shukla, A. K. *et al.* Deficiency of phyto-available sulphur, zinc, boron, iron, copper and manganese in soils of India. *Sci. Rep.* **11**, 19760. <https://doi.org/10.1038/s41598-021-99040-2> (2021).
47. SAS Institute Inc. *SAS/STAT® 15.1 User's Guide* (SAS Institute Inc, Cary, 2018).
48. Falconer, D. S. *Introduction to Quantitative Genetics* 3rd edn. (Longman, New York, 1989).
49. Elshire, R. J. *et al.* A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**(5), e19379. <https://doi.org/10.1371/journal.pone.0019379> (2011).
50. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btm308> (2007).
51. McCormick, R. F. *et al.* The Sorghum bicolor reference genome: Improved assembly and annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* **93**(2), 338–354. <https://doi.org/10.1111/tpj.13781> (2017).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
53. Ali, M., Zhang, Y., Rasheed, A., Wang, J. & Zhang, L. Genomic prediction for grain yield and yield-related traits in Chinese winter wheat. *Int. J. Mol. Sci.* **21**(4) 1342. <https://doi.org/10.3390/ijms21041342> (2020).
54. Wang, J. & Zhang, Z. GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Geno. Proteom. Bioinform.* **19**(4), 629–640. <https://doi.org/10.1016/j.gpb.2021.08.005> (2021).
55. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290. <https://doi.org/10.1093/bioinformatics/btg412> (2004).
56. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucl. Acids Res.* <https://doi.org/10.1093/nar/gkz239> (2019).
57. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* <https://doi.org/10.1101/gr.094052.109> (2009).
58. RStudio Team. *RStudio: Integrated Development for R*. RStudio (PBC, Boston, MA, 2020). <http://www.rstudio.com/>
59. Kuo, K. H. M. Multiple testing in the context of gene discovery in sickle cell disease using genome-wide association studies. *Genom. Insights* <https://doi.org/10.1177/1178631017721178> (2017).
60. Xu, Y. *et al.* Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant Sci.* **9**, 1311. <https://doi.org/10.3389/fpls.2018.01311> (2018).
61. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186. <https://doi.org/10.1093/nar/gkr944> (2012).
62. Sreedasyam, A. *et al.* JGI Plant Gene Atlas: An updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkad616> (2023).
63. Chen, C. *et al.* TBtools-II: A “One for All, All for One” bioinformatics platform for biological big-data mining. *Molec. Plant.* <https://doi.org/10.1016/j.molp.2023.09.010> (2023).
64. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52**(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591> (1965).
65. Ryan, T. A. & Joiner, B. L. *Normal Probability Plots and Tests for Normality* (Technical Report, Statistics Department, The Pennsylvania State University, 1976).
66. Chouraki, V. & Seshadri, S. Genetics of Alzheimer’s Disease. In *Advances in Genetics* (eds by Friedmann, T., Dunlap, J. C., & Goodwin, S. F.) 245–294. <https://doi.org/10.1016/b978-0-12-800149-3.00005-6> (2014).
67. Alemu, A. *et al.* Multi-model genome-wide association and genomic prediction analysis of 16 agronomic, physiological and quality related traits in ICARDA spring wheat. *Euphytica* **217**, 205. <https://doi.org/10.1007/s10681-021-02933-6> (2021).
68. Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaepller, S. M. & Babu, R. Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theor. Appl. Genet.* **128**, 851–864. <https://doi.org/10.1007/s00122-015-2475-3> (2015).
69. Bhatta, M. *et al.* Genome-wide association study reveals novel genomic regions associated with 10 grain minerals in synthetic hexaploid wheat. *Int. J. Mol. Sci.* **19**(10), 3237. <https://doi.org/10.3390/ijms19103237> (2018).
70. Özkuru, E. *et al.* Genome-wide association studies of molybdenum and selenium concentrations in *C. arietinum* and *C. reticulatum* seeds. *Mol. Breed.*, 39(3). <https://doi.org/10.1007/s11032-019-0947-x> (2019).
71. Gunjača, J. *et al.* Genome-wide association studies of mineral content in common bean. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2021.636484> (2021).
72. Lee, Y. *et al.* Genome-wide association study (GWAS) of the agronomic traits and phenolic content in sorghum (*Sorghum bicolor* L.) genotypes. *Agronomy* **13**(6), 1449. <https://doi.org/10.3390/agronomy13061449> (2023).
73. Kawahigashi, H., Kasuga, S., Okuizumi, H., Hiradate, S. & Yonemaru, J. Evaluation of Brix and sugar content in stem juice from sorghum varieties. *Grassl. Sci.* **59**(1), 11–19. <https://doi.org/10.1111/grs.12006> (2013).
74. Wang, J., Yu, J., Lipka, A. E. & Zhang, Z. Interpretation of Manhattan plots and other outputs of genome-wide association studies. In *Genome-Wide Association Studies. Methods in Molecular Biology* Vol. 2481 (eds Torkamaneh, D. & Belzile, F.) (Humana, New York, 2022). [https://doi.org/10.1007/978-1-0716-2237-7\\_5](https://doi.org/10.1007/978-1-0716-2237-7_5).
75. Yoosefzadeh-Najafabadi, M., Eskandari, M., Belzile, F. & Torkamaneh, D. Genome-wide association study statistical models: A review. In *Genome-Wide Association Studies Methods in Molecular Biology* Vol. 2481 (eds Torkamaneh, D. & Belzile, F.) (Humana, New York, 2022). [https://doi.org/10.1007/978-1-0716-2237-7\\_4](https://doi.org/10.1007/978-1-0716-2237-7_4).
76. Abed, A. & Kehel, Z. Preparation and curation of multiyear, multilocation, multitrait datasets. In *Genome-Wide Association Studies. Methods in Molecular Biology* Vol. 2481 (eds Torkamaneh, D. & Belzile, F.) (Humana, New York, 2022). [https://doi.org/10.1007/978-1-0716-2237-7\\_6](https://doi.org/10.1007/978-1-0716-2237-7_6).
77. Ashok Kumar, A., Anuradha, K. & Ramaiah, B. Increasing grain Fe and Zn concentration in sorghum: progress and way forward. *J. SAT Agric. Res.* **11**, 1–5 (2013).

78. Reddy, B. V. S., Ramesh, S. & Longvah, T. Prospects of breeding for micronutrients and  $\beta$ -carotene-dense sorghums. *Intern. Sorghum Millets Newsl.* **46**, 10–14 (2005).
79. Reddy, P. S. *et al.* Association of grain Fe and Zn contents with agronomic traits in sorghum. *Indian J. Plant Genet. Res.* **23**, 280–284 (2010).
80. Nguni, D., Geleta, M., Hofvander, P., Fatih, M. & Bryngelsson, T. Comparative genetic diversity and nutritional quality variation among some important Southern African sorghum accessions [*Sorghum bicolor* (L.) Moench]. *Aust. J. Crop Sci.* **6**, 56–64 (2012).
81. Velu, G. *et al.* Prospects of breeding biofortified pearl millet with high grain iron and zinc content. *Plant Breed.* **126**, 182–185. <https://doi.org/10.1111/j.1439-0523.2007.01322.x> (2007).
82. Velu, G., Rai, K. N., Sahrawat, K. L. & Sumalini, K. Variability for grain iron and zinc contents in pearl millet hybrids. *J. SAT Agric. Res.* **6**, 1–4 (2008).
83. Gupta, S. K., Velu, G., Rai, K. N. & Sumalini, K. Association of grain iron and zinc content with grain yield and other traits in pearl millet (*Pennisetum glaucum* (L.) R. Br.). *Crop Improv.* **36**, 4–7 (2009).
84. Rai, K. N., Govindaraj, M. & Rao, A. S. Genetic enhancement of grain iron and zinc content in pearl millet. *Qual. Assur. Saf. Crops Foods.* **4**, 119–125. <https://doi.org/10.1111/j.1757-837X.2012.00135.x> (2012).
85. Govindaraj, M. *et al.* Combining ability and heterosis for grain iron and zinc densities in pearl millet. *Crop Sci.* **53**, 507–517. <https://doi.org/10.2135/cropsci2012.08.0477> (2013).
86. Kanatti, A. *et al.* Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. *SpringerPlus* **3**, 763. <https://doi.org/10.1186/2193-1801-3-763> (2014).
87. Oikeh, S. O., Menkir, A., Maziya-Dixon, B., Welch, R. & Glahn, R. P. Assessment of concentrations of iron and zinc and bioavailable iron in grains of early maturing tropical maize varieties. *J. Agric. Food Chem.* **51**, 3688–3694. <https://doi.org/10.1021/jf0261708> (2003).
88. Oikeh, S. O. *et al.* Environmental stability of iron and zinc concentrations in grain of elite early-maturing tropical maize genotypes grown under field conditions. *J. Agric. Sci.* **142**, 543–551. <https://doi.org/10.1017/S0021859604004733> (2004).
89. Stangoulis, J. C. R., Huynh, B. L., Welch, R. M., Choi, E. Y. & Graham, R. D. Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. *Euphytica* **154**, 289–294. <https://doi.org/10.1007/s10681-006-9211-7> (2007).
90. Anandan, A., Rajiv, G., Eswaran, R. & Prakash, M. Genotypic variation and relationships between quality traits and trace elements in traditional and improved rice (*Oryza sativa* L.) genotypes. *J. Food Sci.* **76**(4), H122–H130. <https://doi.org/10.1111/j.1750-3841.2011.02135.x> (2011).
91. Garvin, D. F., Welch, R. M. & Finley, J. W. Historical shifts in the seed mineral micronutrient concentration of US hard red winter wheat germplasm. *J. Sci. Food Agric.* **86**, 2213–2220. <https://doi.org/10.1002/jsfa.2601> (2006).
92. Peleg, Z. *et al.* Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat  $\times$  wild emmer wheat RIL population. *Theor. Appl. Genet.* **119**, 353–369. <https://doi.org/10.1007/s00122-009-1044-z> (2009).
93. Zhang, Y. *et al.* Mineral element concentrations in grains of Chinese wheat cultivars. *Euphytica* **174**, 303–313. <https://doi.org/10.1007/s10681-009-0082-6> (2010).
94. Velu, G., Ortiz-Monasterio, I., Singh, R. P. & Payne, T. Variation for grain micronutrients in wheat core collections accession of diverse origin. *Asian J. Crop Sci.* **3**, 43–48. <https://doi.org/10.3923/ajcs.2011.43.48> (2011).
95. Yu, K. *et al.* Genome-wide association studies reveal novel QTLs, QTL-by-environment interactions and their candidate genes for tocopherol content in soybean seed. *Front. Plant Sci.* **13**, 1026581. <https://doi.org/10.3389/fpls.2022.1026581> (2022).
96. Zhang, Y. M., Jia, Z. & Dunwell, J. M. Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Front. Plant Sci.* **10**, 100. <https://doi.org/10.3389/fpls.2019.00100> (2019).
97. Chang, M., He, L. & Cai, L. An overview of genome-wide association studies. In *Computational Systems Biology Methods in Molecular Biology* Vol. 1754 (ed. Huang, T.) (Humana Press, New York, 2018). [https://doi.org/10.1007/978-1-4939-7717-8\\_6](https://doi.org/10.1007/978-1-4939-7717-8_6).
98. He, L. *et al.* Evaluation of genomic prediction for pasmo resistance in flax. *Int. Mol. Sci.* **20**, E359. <https://doi.org/10.3390/ijms20020359> (2019).
99. Li, M. *et al.* Enrichment of statistical power for genome-wide association studies. *BMC Biol.* **12**, 73. <https://doi.org/10.1186/s12915-014-0073-5> (2014).
100. Li, C. X. *et al.* Molecular marker assisted breeding and genome composition analysis of Zhengmai 7698, an elite winter wheat cultivar. *Sci. Rep.* **8**, 322. <https://doi.org/10.1038/s41598-017-18726-8> (2018).
101. Sul, J. H., Martin, L. S. & Eskin, E. Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genet.* <https://doi.org/10.1371/journal.pgen.1007309> (2018).
102. Yuan, Y. *et al.* Genome-wide association mapping and genomic prediction analyses reveal the genetic architecture of grain yield and flowering time under drought and heat stress conditions in maize. *Front. Plant Sci.* **9**, 1919. <https://doi.org/10.3389/fpls.2018.01919> (2019).
103. Chakraborty, P., Biswas, A., Dey, S., Bhattacharjee, T. & Chakrabarty, S. Cytochrome P450 gene families: Role in plant secondary metabolites production and plant defense. *J. Xenobiot.* **13**(3), 402–423. <https://doi.org/10.3390/jox13030026> (2023).
104. Li, L., Chang, Z., Pan, Z., Fu, Z. & Wang, X. Modes of heme binding and substrate access for cytochrome P450 CYP74A revealed by crystal structures of allene oxide synthase. *Proc. Nat. Acad. Sci. U. S. A.* **105**(37), 13883–13888. <https://doi.org/10.1073/pnas.0804099105> (2008).
105. Sun, K. *et al.* Genome-wide analysis of the cytochrome P450 gene family involved in salt tolerance in *Gossypium hirsutum*. *Front. Plant Sci.* **12**, 685054. <https://doi.org/10.3389/fpls.2021.685054> (2021).
106. Satyavathi, C. T. *et al.* Stage specific comparative transcriptomic analysis to reveal gene networks regulating iron and zinc content in pearl millet [*Pennisetum glaucum* (L.) R. Br.]. *Sci. Rep.* **12**, 276. <https://doi.org/10.1038/s41598-021-04388-0> (2022).
107. Roda, F. A. *et al.* Rice biofortification with zinc and selenium: A transcriptomic approach to understand mineral accumulation in flag leaves. *Front. Genet.* **11**, 543. <https://doi.org/10.3389/fgene.2020.00543> (2020).
108. Hochmuth, G. Iron (Fe) nutrition of plants: SL353/SS555, 8/2011. EDIS 2011 (8) <https://doi.org/10.32473/edis-ss555-2011> (2011).
109. Araki, R., Takano, Y., Miyazaki, H., Ii, H. & An, P. Drought stress alters iron accumulation in *Sorghum bicolor* seeds. *Environ. Exp. Bot.* **204**, 105093. <https://doi.org/10.1016/j.envexpbot.2022.105093> (2022).
110. Shabbir, Z. *et al.* Copper uptake, essentiality, toxicity, detoxification and risk assessment in soil-plant environment. *Chemosphere* **259**, 127436. <https://doi.org/10.1016/j.chemosphere.2020.127436> (2020).
111. Varshney, R. K. *et al.* Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant Sci.* **26**(6), 631–649. <https://doi.org/10.1016/j.tplants.2021.03.010> (2021).
112. Sehgal, D. *et al.* Haplotype-based, genome-wide association study reveals stable genomic regions for grain yield in CIMMYT spring bread wheat. *Front. Gen.* <https://doi.org/10.3389/fgene.2020.589490> (2020).
113. Bhat, J. A. *et al.* Identification of superior haplotypes in a diverse natural population for breeding desirable plant height in soybean. *Theor. Appl. Genet.* **135**(7), 2407–2422. <https://doi.org/10.1007/s00122-022-04120-0> (2022).
114. Meuwissen, T. H., Odegard, J., Andersen-Ranberg, I. & Grindflek, E. On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genet. Sel. Evol.* **46**, 1–8. <https://doi.org/10.1186/1297-9686-46-49> (2014).
115. Zhang, L. M. *et al.* Sweet sorghum originated through selection of *Dry*, a plant-specific NAC transcription factor gene. *Plant Cell.* **30**(10), 2286–2307. <https://doi.org/10.1105/tpc.18.00313> (2018).

116. Lin, Z. *et al.* Parallel domestication of the *Shattering1* genes in cereals. *Nat Genet* **44**(6), 720–724. <https://doi.org/10.1038/ng.2281> (2012).
117. Wu, X. *et al.* Genomic footprints of sorghum domestication and breeding selection for multiple end uses. *Mol. Plant* **15**(3), 537–551. <https://doi.org/10.1016/j.molp.2022.01.002> (2022).
118. Galotta, M. F., & Roberts, I. N. A method for in vivo determination of subtilase activity in germinating seeds. *J. Cereal Sci.*, 85, 105–110 <https://doi.org/10.1016/j.jcs.2018.11.014> (2019).
119. Schaller, A. *et al.* From structure to function—a family portrait of plant subtilases. *New Phytol.* **218**(3), 901–915. <https://doi.org/10.1111/nph.14582> (2017).
120. Lance, C. & Rustin, P. The central role of malate in plant metabolism. *Physiol. Vég.* **22**, 625–641 (1984).
121. Martinoia, E. & Rentsch, D. Malate compartmentation—responses to a complex metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 447–467 (1994).
122. Gietl, C. Malate dehydrogenase isoenzymes: Cellular locations and role in the flow of metabolites between cytoplasm and cell organelles. *Biochem. Biophys. Acta* **1100**, 217–234 (1992).
123. Kochian, L. V. Cellular mechanisms of aluminum toxicity and resistance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 237–260 (1995).
124. Ebbisa, A. Mechanisms underlying cereal/legume intercropping as nature-based biofortification: A review. *Food Prod. Process. Nutr.* **4**, 19. <https://doi.org/10.1186/s43014-022-00096-y> (2022).

## Acknowledgements

Niranjan Ravindra Thakur acknowledges the funding support for this research from the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India (ICRISAT). The authors thank Dr. Kamalakar Kamble and Mr. Jayakumar Jaganathan for their help in executing and managing the trials, and phenotypic data curation.

## Author contributions

N.R.T., S.P.M., H.V.K., and P.R.S. were involved in performing the experiments. A.V. analyzed the data. P.R. and N.R.T. involved in GWAS analysis and images and visualization. P.R. and D.A.O. provided analysis tools. P.R. conducted bioinformatics to identify associated candidate genes. N.R.T. and S.G. wrote the first draft of the manuscript, and D.A.O. and E.H. edited and revised the manuscript. E.H. conceived, designed, supervised the experiments, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-63308-0>.

**Correspondence** and requests for materials should be addressed to E.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024