RESEARCH ARTICLE

Plants People Planet PPP
Open Access

Genome-wide association mapping and comparative genomics identifies genomic regions governing grain nutritional traits in finger millet (*Eleusine coracana* L. Gaertn.)

Swati Puranik[1,2] (iD) | **Pranav P. Sahu**[1,2] (iD) | **Sarah Beynon**[1] | **Rakesh K. Srivastava**[3] (iD) | **Deepmala Sehgal**[4] (iD) | **Henry Ojulong**[5] | **Rattan Yadav**[1] (iD)

[1]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK

[2]Global Change Research Institute of Czech Academy of Sciences, Brno, Czech Republic

[3]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India

[4]International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

[5]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Nairobi, Kenya

Correspondence
Rattan Yadav, Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK.
Email: rsy@aber.ac.uk

Societal Impact Statement

Micronutrient deficiency is a serious and underestimated global health concern. Identifying existing micronutritional richness in traditional crops, and breeding this potential into staple crops that are more frequently consumed, could offer a potential low-cost, sustainable solution to micronutrient deficiency. Here, we provide the first insight into genetic control of grain micronutrient content in the staple food crop finger millet (*Eleusine coracana*). Quantifying the existing natural variation in nutritional traits, and identifying the regions of the genome associated with these traits, will underpin future breeding efforts to improve not only global food and nutrition security, but also human health.

Summary

- Finger millet is an excellent cereal crop to alleviate micronutrient malnutrition in marginal communities because of its nutrient-dense grains. Identification of the alleles governing these characteristics will help to develop improved germplasm.

- An assembly of 190 genotypes was evaluated for six minerals (iron, zinc, calcium, magnesium, potassium, and sodium) and protein content. A combination of genotyping-by-sequencing (GBS) and genome-wide association study (GWAS) was applied to identify marker-trait associations (MTAs). Candidate genes underlying significant associations were predicted through comparative mapping with other cereals.

- A wide range of variation was observed for all traits. GBS generated 169,365 single-nucleotide polymorphisms and three subpopulations were identified. GWAS, using general linear and mixed model approaches to correct for population structure and genetic relatedness, identified 418 common markers (p-value ≤ 10^{-3}, FDR < 0.1) linked with mineral content. Of these, 34 markers crossed the Bonferroni threshold, out of which 18 showed homology with candidate genes having putative functions in binding, remobilization or transport of metal ions.

- This is the first report to utilize the phenotypic variability of grain minerals in finger millet genotypes to identify MTAs and predict associated putative candidate genes. Postvalidation, these markers may be employed to improve grain nutrient quality through marker-assisted breeding.

1 | INTRODUCTION

Key micronutrients such as iron, zinc, and calcium are essential to almost every process in living organisms. Dietary deficiency, also known as micronutrient (MN) malnutrition or hidden hunger, impacts almost 2 billion people worldwide across all age groups, genders, and ethnicities (FAO & International Life Science Institute, 1997). Iron and zinc deficiencies affect almost 60% and 30% of the global population, respectively, and MN deficiency thus is a major global health concern to both low- and high-income countries (McLean, Cogswell, Egli, Wojdyla, & De Benoist, 2009). Where populations are weakened due to hidden hunger, individuals are at increased risk of succumbing to infection or developing chronic disease. As such, these account for almost 80% of deaths in malnourished populations. Young children, pregnant, and nursing women in the poorest regions of the world are at the highest risk of suffering from such malnutrition. In developing countries, these groups are also the least likely to be able to afford necessary and timely treatment for conditions related to MN malnutrition placing considerable pressure on both the healthcare sector and family income. Increasing the micronutritional value of staple crops through targeted breeding programs involving varieties with the highest nutritional contents may be a sustainable low-cost solution.

Millets are a group of traditional and heterogeneous cereals often grown in harshest areas of Asia and Africa (Ramakrishnan et al., 2017). Recently, interest in this group has increased and they have gained much attention as potential crops for a "New Green Revolution" due to their high inherent nutritional quality and environmental hardiness (Goron & Raizada, 2015; Padulosi et al., 2009). Finger millet (*Eleusine coracana* L. Gaertn.), an allotetraploid (2n = 4x = 36, AABB) annual millet, is predominantly grown across Asia and Africa and contributes to 12% of the global cultivated area under millets (Vetriventhan, Upadhyaya, Dwivedi, Pattanashetti, & Singh, 2015). It is the fourth most important millet after sorghum, pearl millet, and foxtail millet (Upadhyaya, Gowda, & Reddy, 2007). In hot, arid regions, with low soil fertility, it is able to produce reasonable grain and fodder yields. This ability can partly be attributed to its efficient carbon concentrating mechanism, the C4 pathway (Hittalmani et al., 2017). It is also a rich source of several essential amino acids and health benefitting MNs, phytochemicals, and vitamins (Puranik et al., 2017). Amino acids like lysine and methionine are often scarce in plant food crops, but they are found in abundance in finger millet. When compared to other cereals, finger millet also contains high concentration

of calcium (350 mg/100 g) in its grains and can be an inexpensive food to treat problems related to osteoporosis (Puranik et al., 2017). Alongside calcium, finger millet grains are known to contain a high quantity of other MNs including iron, zinc, phosphorus, and potassium (Shashi, Sharan, Hittalamani, Shankar, & Nagarathna, 2007; Tripathi & Plate l, 2010). Other important properties of finger millet are that it is gluten-free (helpful to patients suffering from celiac disease), has low glycemic index (keeps a control on blood sugar level), and possesses excellent malting (nourishing food for infants) and nutraceutical properties (Kumar, Metwal, et al., 2016).

Major breeding targets for finger millet have included traits such as improvement of agro-morphological characteristics, calcium accumulation, blast resistance, nitrogen use efficiency, and tolerance to phosphorus deficiency (reviewed extensively in Sood et al., 2016 and Gupta et al., 2017). Utilization of molecular marker technology is now emerging as an important tool in selection and breeding programs including traits that are expensive to phenotype and have complex genetic architecture. Important genes or genomic regions underlying complex quantitative traits can be tagged, by markers using biparental or association mapping approaches (Sehgal et al., 2015). Mapping using biparental mapping-population has limitations of low number of segregating alleles and lower mapping resolution (Sood et al., 2016). Association mapping (AM), on the other hand, utilizes the existing genetic diversity in natural germplasm populations and exploit their historic recombination events to map or fine map genes. This approach, also known as linkage disequilibrium (LD) mapping, has a higher mapping resolution due to the use of a genetically diverse population (Buckler & Thornsberry, 2002). AM has been used in Arabidopsis (Kooke et al., 2016), rice (Agrama, Eizenga, & Yan, 2007), soybean (Li, Zhao, Han, Li, & Xie, 2018), tomato (Albert et al., 2016), maize (Yu & Buckler, 2006), sorghum (Kulwal, 2016), foxtail millet (Gupta, Kumari, Muthamilarasan, Parida, & Prasad, 2014), and pearl millet (Sehgal et al., 2015) for identifying regions associated with diverse traits including yield components, morphology, stress tolerance, and seed components.

In finger millet, AM has been used to identify quantitative trait loci (QTLs) associated with different agro-morphological traits, protein and tryptophan contents as well as blast tolerance, and low phosphorus tolerance (Babu, Agrawal, Pandey, Jaiswal, & Kumar, 2014; Babu, Agrawal, Pandey, & Kumar, 2014; Babu, Dinesh, et al., 2014; Ramakrishnan, Ceasar, Duraipandiyan, Vinod, et al., 2016; Ramakrishnan et al., 2017). In recent years, Sharma et al. (2018) used

SNP markers to conduct genome-wide association mapping of major agro-morphological traits in finger millet. Compared to these traits, understanding of the genetic basis of nutrient accumulation in finger millet grains remains limited.

Our objectives in this study were to generate a large set of genome-wide markers (single-nucleotide polymorphism; SNP) through genotyping by sequencing (GBS) and to demonstrate their use in capturing genetic variations associated with nutritional traits through genome-wide association studies (GWAS). We made use of a population of 190 genotypes assembled by combining individuals from core, minicore, and elite varieties for generating SNP variants, and phenotyped them for grain minerals such as iron, zinc, calcium, potassium, sodium, magnesium, and for total protein content.

2 | MATERIALS AND METHODS

2.1 | Plant material and assembly of GWAS population

A GWAS population was developed using a set of 190 accessions of finger millet with diverse geographic origins (Table S1). This experimental population included 142 traditional cultivar/landrace accessions, which are derived from the finger millet core and minicore collection at ICRISAT representing the entire trait diversity in finger millet germplasm (Upadhyaya, Gowda, Pundir, Reddy, & Singh, 2006; Upadhyaya et al., 2010). These accessions have their origins across Africa (Zimbabwe, Kenya, Uganda, Malawi, Zambia, Nigeria, Ethiopia, Tanzania, Burundi, and Senegal), Asia (India, Nepal, Sri Lanka, Pakistan, and Maldives), Europe (Italy, United Kingdom, and Germany), and North America (United States of America and Mexico). The GWAS population also consisted of 48 genotypes representing elite cultivars that are preferred by the local farmers in the countries of Kenya, Tanzania, and Uganda of the East African region.

The population was grown at ICRISAT-Kiboko Station, Kenya during the long rainy season between April and July 2015. The station is located at latitudes 2.2172°S and longitudes 37.72°E and at 975 m above sea level with acri-rhodic ferralsols soils. The crop was raised in an augmented incomplete design with 20 blocks, each block consisting of 32 accessions and 2 check varieties namely U15 and P224 replicated twice per block. Each entry was grown in single rows plots of 4-m length with inter row and intra row spacing of 50 cm and 10 cm, respectively. The grains from the individual entries of the population were harvested for mineral and protein analysis. Best linear unbiased predictions (BLUPs) of the entries were calculated, thereby adjusting the influence of the neighboring rows. These BLUPs were used downstream analysis.

2.2 | DNA extraction and genotyping

The population was grown in a greenhouse setting at IBERS, Aberystwyth University, UK. About 100 mg leaf tissue was harvested from 3-week-old seedlings and immediately frozen in liquid nitrogen. The samples were ground to a fine powder using TissueLyser II (QIAGEN). DNeasy 96 Plant Kit (QIAGEN) was used to extract genomic DNA, which was then quantified by Qubit 2.0 fluorometer (Invitrogen) and agarose (0.8%; w/v) gel electrophoresis. The quality of DNA was further ascertained by *Eco*RI-based restriction digestion of 20 samples (representing 10% of the genotypes). Samples were normalized to a final concentration of 30 ng/ul and sent to the Genomic Diversity Facility at Cornell University (USA). Two *Ape*KI-based genotyping-by sequencing (GBS) libraries, each with 95 genotypes and one random blank control, were prepared following the protocol of Elshire et al. (2011) and named as CaFMillet1 and CaFMillet2. Four lanes of single-end Illumina HiSeq 2,500 were used to sequence two 96-plex libraries (CaFMillet_Pl1 & CaFMillet_Pl2).

2.3 | SNP identification and filtering

The raw sequence files were parsed based on their barcodes and all reads were trimmed to 64 bps. In the absence of a finger millet reference genome sequence, SNPs were called *de novo* using the UNEAK pipeline (Lu et al., 2013). SNPs were also called using the GBS pipeline (Glaubitz et al., 2014) and the reference genome sequences of cereals closely related to finger millet such as tef (*Eragrostis tef*), foxtail millet (*Setaria italica*) and rice (*Oryza sativa*). Both pipelines were implemented in the TASSEL 3.0 package (Bradbury et al., 2007). All the pipeline arguments used are listed in Table S2.

Genotype likelihood scores were calculated based on Etter, Bassham, Hohenlohe, Johnson, and Cresko (2011) and the most probable genotype was assigned as a function of a genotype quality (GQ) score (https://software.broadinstitute.org/gatk/ calculated according to the GATK version; https://software.broadinstitute.org/gatk/). VCFtools version v0.1.12a was used to calculate the depth and missingness (Danecek et al., 2011). The remaining SNPs obtained from both of the pipelines were filtered according to commands described in Table 1. Initially, a set of SNPs that passed stringent filtering criteria (i.e., "stringent" filtering, Table 1) was generated to validate data quality as well as population structure analysis. Furthermore, larger SNP sets were generated using "relaxed" filtering criteria for GWAS analyses (Table 1). We used the PLINK software tool (Purcell et al., 2007) to calculate the number of unlinked single-nucleotide variants.

2.4 | Population structure and genetic relatedness analysis

Discriminant analysis of principal components (DAPC) (Jombart, Devillard, & Balloux, 2010) was employed to analyze the population genetic structure. The most likely number of clusters were inferred using the R package Adgenet (Jombart, 2008; Jombart & Ahmed, 2011; R Core Team, 2013) and its find.clusters function. The appropriate principle components were calculated from the probability of assignment

TABLE 1 Filtering criteria for GBS generated SNP data

Filtering Parameter	Tool	Commands for relaxed filtering	Commands for stringent filtering
• Sites by number of alleles	VCF TOOLS	--min-alleles 2 --max-alleles 2	--min-alleles 2 --max-alleles 2
• Sites by allele frequency and call rate		--maf 0.01	--maf 0.01
• Sites by allele count and missing data		--mac 3	--mac 10
• Individuals by mean depth		--min-indv-meanDP 2	--min-indv-meanDP 2
• Minimum site coverage	TASSEL GUI 5.2.31	152 out of 191 sequences	152 out of 191 sequences
• Minimum frequency		0.01	0.05
• Remove minor SNP states		TRUE	TRUE
• Minimum taxon coverage		0.1	0.1
• WINDOW 1,000, LD 0.5	PLINK	-	>./plink --bfile binaryfilename --make-founders --indep-pairwise 1,000 50 0.5 --out givenewname
			>./plink --bfile binaryfilename --extract givenewname.prune.in --make-bed --out givenewname
			>./plink --bfile givenewname --recode --out givenewname2

of individuals to individual clusters as advised in the manual. Bayesian Information Criterion (BIC) for $K = 1 – 10$ (K = number of populations) was used to indicate the optimal number of populations with minimum observed BIC value. The optimal numbers of PCs were retained through optimization of α-score, which measures the difference between the proportion of successful reassignment of the analysis (observed discrimination) and values obtained using random groups (random discrimination; Jombart, 2008; Jombart & Ahmed, 2011).

The SNP-based genetic groups were employed to estimate genetic distances using Nei's standard genetic distance (Saitou & Nei, 1987). These values were then used to construct a phylogenetic tree using a neighbor-joining (NJ) method in the R "ape" package (Paradis, Claude, & Strimmer, 2004). The "dist" function of the package was employed to calculate Euclidean distance matrix.

2.5 | Analytical determination of mineral and protein content

About 5 g seeds from each entry were finely ground in a Retsch® miller (model ZM 200 GmbH Germany) fitted with a 1-mm-filter mesh. The milled samples were used for determination of six minerals: iron, sodium, potassium, magnesium, calcium, and zinc along with nitrogen for estimating total protein. Briefly, 1 g powdered sample was subjected to overnight aqua regia digestion in 100 ml Kjeldahl flasks. The digest was filtered using filter paper disks (Whatman®) into a 50 ml volumetric flask and further diluted with additional dilute aqua regia mix. The multielement analysis was carried out on each grain digest using inductively coupled plasma optical emission spectroscopy (ICP OES; Optima 8000DV, PerkinElmer, USA) in triplicates. Nitrogen content in the ground seed samples was detected by combustion followed

by thermal conductivity using the Leco FP-528 Nitrogen/Protein Determinator (LECO, 2016). The total nitrogen percentage was multiplied by 6.25 to calculate crude protein content in the grains (Mariotti, Tomé, & Mirand, 2008). In order to ensure experimental accuracy, two standard analytical quality control samples and a blank were included for each run. All the estimations were conducted at IBERS analytical chemistry and metabolomics facility according to the standard association of analytical communities' (AOAC) protocols (AOAC, 2016). Two genotypes were excluded from downstream analysis as they failed the ICP-OES analysis.

2.6 | Linkage disequilibrium and genome-wide association analysis (GWAS)

The stringent filtered set of SNPS was used to conduct analysis of linkage disequilibrium (LD) using TASSEL software with the default settings. The Pearson correlation values (R^2) and pairwise distance between SNPs from above analysis were then imported in R to generate the genome-wide LD decay plots.

In order to conduct GWAS, three statistical models were employed in the software TASSEL 5.0 (Bradbury et al., 2007; Buckler et al., 2009). These included: (a) naïve model with 10,000 permutations (b) general linear model (GLM) including a Q-matrix with 10,000 permutations, and (c) mixed linear model (MLM) with Q-matrix and K-matrix. The two matrices accounted for corrections in population structure (Q) and/or genetic relatedness (K; Dhanapal & Crisosto, 2013; Pasam et al., 2012; Yang et al., 2010; Yu et al., 2006). Significant marker-trait associations (MTAs) were defined based on significance threshold. For GLM, these were kept as − log10 p ≥ 3.00; p ≤ .001 and for MLM as − log10 p ≥ 2.00; p ≤ .01 as described previously by researchers (Hao, Chao, Yin, & Yu, 2012; Yang et al., 2010). For both models, the significance of MTAs

was estimated using a multiple testing approach with Bonferroni correction as well as by the false discovery rate (FDR) method (Benjamini & Hochberg, 1995) employed through the QVALUE package in R (Storey & Tibshirani, 2003). Overall, an MTA was called significant if it had a − log10 $p \geq 3.00$; $p \leq .001$ (for GLM) and − log10 $p \geq 2.00$; $p \leq .01$ (for MLM) and presented an FDR < 0.1 across both the models. The corresponding R^2 values were used to represent proportion of the phenotypic variation explained (PVE) by each marker. The p-values from the models were used as an input file to generate Manhattan and quantile–quantile (QQ) plots using the R package qqman (Turner, 2018).

Later, the tags on physical maps (TOPM) file was opened in TASSEL to retrieve 64 bp sequences containing the SNPs. Due to the lack of a finger millet whole genome sequence, it was difficult to directly estimate the genomic location of identified associations. We therefore used an in silico comparative mapping approach to identify candidate genes based on syntenous relationships among plants. The extracted sequences were searched for sequence similarity against NCBI BLAST, Ensembl Plants server, or Phytozome databases using default parameters to identify putative candidate genes near the QTL sequences (Altschul, Gish, Miller, Myers, & Lipman, 1990; Goodstein et al., 2012; Howe et al., 2020). Hits were considered to be significant on the basis of E value ≤ 0.01. Possible biological roles of the closest hits identified through BLAST were further analyzed for their significance to grain nutrient accumulation.

3 | RESULTS

3.1 | Genotyping and detection of SNP variants

The two libraries (four lanes) generated a total of ~ 66GB of raw data with 1.02×10^9 reads (498,678,076 for CaFMillet_Pl1 and 520,458,445 CaFMillet_Pl2) corresponding to an average of 0.54 million reads per genotype. For library CaFMillet_Pl1, there were 464,622,290 reads (93.2%) with a good barcode and cut site overhang while for CaFMillet_Pl2, it was 94.6% with 492,547,602 reads with an overall average of 93.9% reads with a good barcode. These have been deposited in the NCBI-SRA database under the accession "SRP100423." After merging, 5,436,304 tags were generated of which 576,204 reciprocal tag pairs (two-tag networks resulting in only bi-allelic loci) were identified and analyzed for SNP calling using the TASSEL-UNEAK pipeline. The UNEAK pipeline generated 169,365 SNPs after merging taxa across all 190 finger millet accessions with a mean site depth of 3.99 and 0.32 missing value. Applying the filtering parameters such as those for missing data, minor allele frequency etc. (Table 1), resulted in the detection of 16,000 putative SNPs in the stringent and 73,419 putative SNPs in the relaxed parameter.

3.2 | Population structure analysis

The population structure of this collection was described without any a priori group assignment. The functions "find.clusters" and

"k-means" algorithm, retained 200 principal components (PCs), that accounted for more than 99% of the variance. Figure 1a shows the percentage of variance explained by the first 10 PCs. The α-score is an optimization procedure (reassignment probability for given populations minus the reassignment probability for random permuted groups) was used to evaluate optimal number of PCs to retain. The optimal number of retained PCs (based on the α-score), minimized the number to only two PCs needed for the assignment analysis (Figure 1b). An elbow curve of BIC values, as a function of k, indicated that optimal number of cluster was 3 (Figure 1c). Using this information (3 clusters from BIC and α-optimized 2 PCs from DAPC), a scatter plot was drawn (Figure 1d). This plot distinguished three separate clusters, corresponding to their geographic origins. The first PC which explained 7.63% genetic variation, separated cluster 2 (mainly Asia) from cluster 1 and 3 (African region; Figure 1d). The second PC explained 5.13% genetic variation and mainly separated these latter two populations into further two groups; East African populations (cluster 1) and those originating from southern Africa (cluster 3; Figure 1d). The probability of membership assignments was 100% for cluster 1 and 2 and 98% for cluster 3. Pairwise Fst values among DAPC clusters ranged from 0.047 (Cluster 1-Cluster 3) to 0.074 (Cluster 2-Cluster 3; Table 2) signifying high genetic differentiation of the subpopulations. In addition, the posterior probability plots that were drawn based on the posterior membership probabilities of each individual to either of the three clusters (Table S3) showed the clear genetic clustering of populations (Figure 1e) based on the GBS-SNP data. This corresponded to the groups defined by the *find.clusters* procedure. Only few individuals exhibited a mixed genetic constitution (Figure 1e), mostly within the subpopulation 3/South African grouping. Within the 48 elite genotypes, the 31 (64.5%) were found to have East African origins, whereas South African (9) and Asian genotypes had almost equal shares (8).

The genetic relatedness among accessions as visualized by the NJ tree were corresponded closely to the DAPC analysis (K = 3; Figure 2). The accessions separated into three main groups. An Asian group was observed, containing genotypes originating in India, Nepal, Pakistan, Sri Lanka, and the Maldives. Samples from the African continent were found to separate into an eastern group, originating from Burundi, Ethiopia, Kenya, Tanzania, and Uganda and a southern group including accessions from Malawi, Zambia, and Zimbabwe. Many lines of European or American origin showed little or no clustering with any particular group. Some samples of unknown geographical origin were found to closely align with one of the three subpopulations identified.

3.3 | Phenotypic variation among accessions

The entire range of six minerals (calcium, iron, zinc, sodium, potassium, magnesium) and total protein content in the finger millet GWAS population is shown in Figure 3 and Table S4. Large variation was observed for all traits measured (Table 3). In the panel used in this
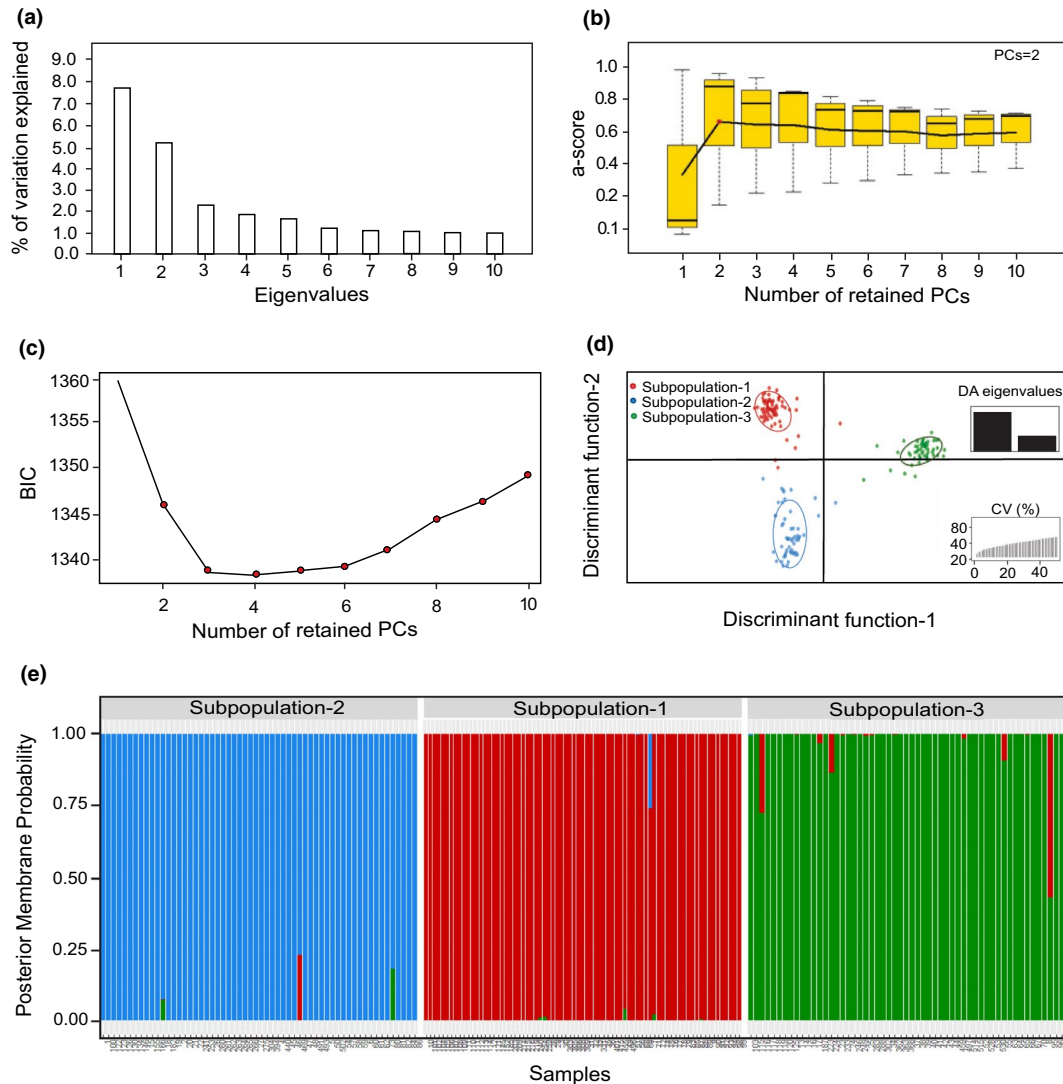
FIGURE 1 Inference of the number of clusters in the DAPC performed on the 16,000 stringently filtered SNP dataset. (a) Cumulative variance (%) explained by the principal component analysis (PCA) relative to the number of PCs retained in the DAPC analysis. (b) Optimization α-score graph. (c) The Bayesian information criterion (BIC) value is plotted is against the numbers of clusters to select optimal genetic populations. The BIC rate of change continuously increases after three clusters, indicating that a K value of 3 (the lowest BIC value) represents the best summary of the data. (d) Scatterplot of the discriminant analysis of principle components (DAPC) using the first 2 PCs distinguishes the three genetic clusters. The clusters are distinguished by colors and 95% inertia ellipses. DA barplot in the inset (top right) displays the proportion of genetic information comprised in each consecutive discriminant function. X and Y axis of the scatterplot describe the first and second discriminant function. (e) Subdivision of the individuals based on the DAPC membership probabilities into three genetic clusters (k* = 3). Same color in different individuals indicates that they belong to the same cluster. Subpopulation 1 (red) is associated with individuals mostly originating from East African countries whereas subpopulation 3 (green) represents those from the south African origin. The Asian origin populations were clustered into the second (blue) cluster. This STRUCTURE-like plot of DAPC analysis also shows some admixed individuals

study, the elite local variety KNE 1,149 had lowest amount of iron in its grains (3.025 mg/100 g), whereas genotype IE2586 had the highest (16.623 mg/100 g). Minicore genotypes IE4816 and IE2957 had the lowest (106.364 mg/100 g) and highest (179.99 mg/100 g) magnesium content, respectively. The range of calcium varied from 223.629 mg/100 g in the elite local variety KNE # 628 to 422.556 mg/100 g in core collection genotype IE6541. Grain potassium value was lowest in IE6059 (266.595 mg/100 g) and highest (668.83 mg/100 g) in IE2589, both minicore genotypes. In core genotype IE5992, 3.86% protein was found,

whereas maximum protein (11.27%) was detected in elite local cultivar accession#32. Similarly, sodium content was only 6.25 mg/100 g in genotype IE2869 but reached a high of 41.44 mg/100 g in the elite variety KNE # 622. About 1.02 mg/100 g zinc was found in the grains of core genotype IE2587, but genotype IE4734 of the minicore collection had 2.66 mg/100 g. Thus, the African elite local varieties were found to have the lowest amount of iron and calcium, but higher amount of protein and sodium content as compared to other minicore/core collection genotypes.

TABLE 2 Pairwise Fst value matrix representing genetic distances among the DAPC identified clusters for the three finger millet subpopulations

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0	0.072173502	0.046692558
Cluster 2	0.072173502	0	0.073910736
Cluster 3	0.046692558	0.073910736	0

The variability across all traits, as estimated by the coefficient of variation, ranged from ~ 8%-37%. The mean of calcium content over all the accessions was 314 mg/100 g which is in agreement with several previous studies (Puranik et al., 2017). Minimum to moderately significant correlations were found between grain MNs and protein content traits (Table 4). Grain iron especially was only weakly correlated with other traits like magnesium and calcium content. Magnesium showed a positive correlation with calcium at a moderate level but a weak correlation with zinc content. Grain potassium content was found to always share a weak negative relation with magnesium and calcium content, but a mild positive correlation with sodium content. Mild correlation between zinc and sodium were weakly negative. Weak positive correlations were also found between sodium and potassium and protein and zinc. None of the traits showed very strong correlations.

When compared within groups, the mineral and protein content were not vastly different (Figure S1). Calcium and potassium content was slightly lower and zinc was higher in the elite local varieties while iron was higher in the minicore accessions (Figure S1). The genotypes belonging to Asian subpopulation had relatively higher calcium content than the African population, whereas that latter were richer in potassium and zinc content (Figure. S2).

3.4 | Linkage disequilibrium, genome-wide association analysis, and identification of candidate genes

As finger millet genome sequencing is still in its infancy, it is difficult to predict the exact rate of LD decay. The rate of LD decay as measured by the R^2 values (squared (Pearson correlation) and plotted against physical distance between markers is shown in Figure S3. Across the whole genome, of 65,536 pairwise combinations obtained from 16,000 filtered marker loci, 324 (~0.5%) pairs had $R^2 \geq 40\%$. About 16.84% of the SNP pairs ($R^2 \geq 0.2$; $p < .05$) were found to be in LD. The maximum R^2 value dropped to its half (0.5) as distance 28.027 kbp.

Using the relaxed filtered set of SNPs (73,419), naive GWAS analyses without incorporating effects of population structure and relatedness showed very inflated p-values (Figure 4a; Figures S4–S9). Inclusion of the first two principle components as correction for existing population structure alleviated this for all traits. The GLM approach identified a total of 2085 SNPs significantly ($p \leq .001$ and
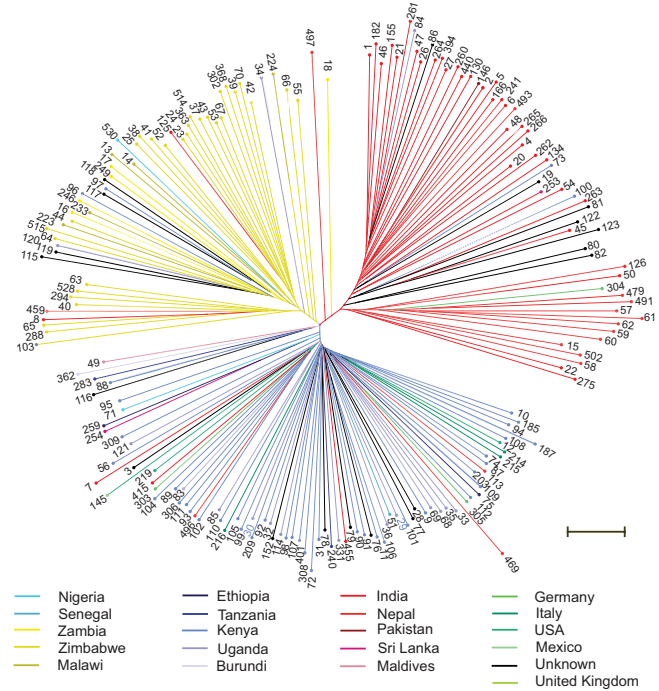

FIGURE 2 Genetic relatedness among the finger millet genotypes. Neighbor-joining method was used to construct the unrooted tree where the colors differ by countries of origin. Each branch represents one accession
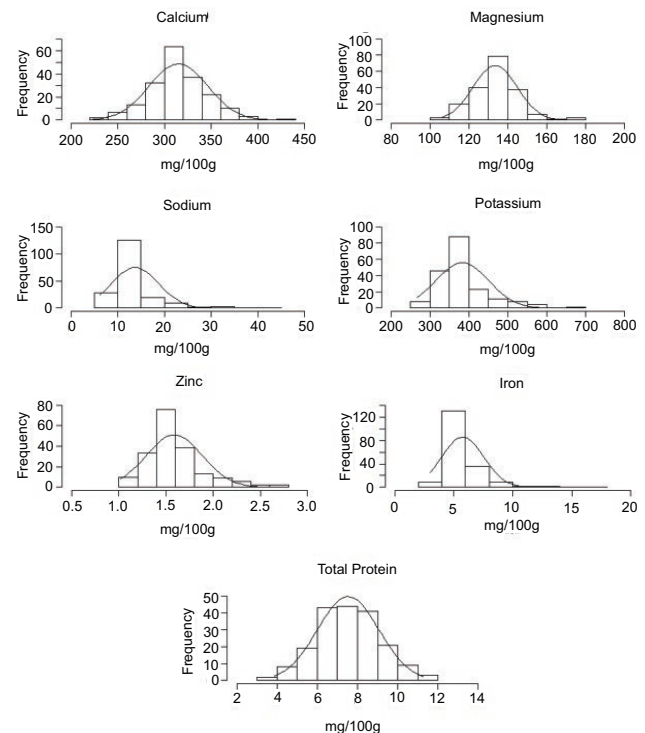
Legend:
- Nigeria
- Senegal
- Zambia
- Zimbabwe
- Malawi
- Ethiopia
- Tanzania
- Kenya
- Uganda
- Burundi
- India
- Nepal
- Pakistan
- Sri Lanka
- Maldives
- Germany
- Italy
- USA
- Mexico
- Unknown
- United Kingdom

FIGURE 3 Histogram representing distribution of phenotypic variation in minerals and protein content, respectively, in the various genotypes as analyzed using ICP-OES and combustion analyses

TABLE 3 Phenotypic variation in traits analyzed in the whole collection

Trait	Min.	Max.	Mean	St. dev.	CV%
Calcium (mg/100g)	223.629	422.556	314.819	31.122	9.886
Iron (mg/100g)	3.025	16.623	5.776	1.766	30.576
Sodium (mg/100g)	6.254	41.441	13.731	5.038	36.689
Potassium (mg/100g)	266.595	668.830	383.156	67.593	17.641
Magnesium (mg/100g)	106.364	179.991	133.543	11.243	8.419
Zinc (mg/100g)	1.021	2.660	1.589	0.297	18.698
Protein (%w/w)	3.859	11.271	7.535	1.523	20.215

FDR < 0.1) associated with the traits (Figure 4b; Figures S4–S9). The addition of the IBS kinship matrix (K) to the GLM analysis further reduced the inflated p-values. Using the conservative MLM approach, we detected a total of 430 putative associations for SNPs (Figure 4c; Figures S4–S9). Overall, using both the models (GLM and MLM) we identified 418 common MTAs (Dataset S1). These analyses also identified several MTAs for calcium and zinc content that were above the set p-value and FDR criteria in at least one of the models (Dataset S1; Figures S7–S8).

For grain iron content in finger millet genotypes, 894 makers were identified using GLM while 148 associations were found using MLM approach (Dataset S1; Figure S4). All the MLM model generated MTAs, except for the marker S1_42935743, were found to associate with MTAs for iron content identified through the GLM analysis. Of the 148 MTAs, a locus at marker S1_5895347 was the most strongly associated with iron and explained 24.59% phenotypic variation. The 444 markers that were identified to be associated with potassium content through GLM analysis were reduced to 174 markers after correcting for family relatedness (K). In both these analysis, 164 markers were found to be in common (Dataset S1; Figure S5). The most significant association was shown by S1_55418346 explaining 18.03% variation. Grain sodium content was another trait with high number of trait-SNPs associations. GLM- and MLM-based association analysis revealed 639 and 106 MTAs, respectively (Dataset S1; Figure S6). One hundred and four (104) MTAs were collectively presented for this trait through both the approaches. About 18.66% phenotypic variation for sodium content was explained by marker S1_47207745. Both GLM and MLM identified five common significant MTA for magnesium content viz. S1_47630040, S1_463458, S1_33226241, S1_37853063, and S1_23369654 (Dataset S1; Figure 4a-4c). Of this, the marker S1_47630040 was associated with magnesium with a R^2% value 18.84%. Although the GLM approach identified 96 markers to be significantly associated with finger millet zinc content, none met the FDR cut-off parameter in the MLM model (Dataset S1; Figure S7). Similarly, GLM model identified 7 MTAs for calcium content above the p-value and FDR threshold. Although these markers were also identified through the MLM approach, they did not cross the FDR < 0.1 threshold (Dataset S1; Figure S8). Both GLM and MLM models identified several SNPs to be significantly ($p \le .001$) associated with grain protein content as well, however, they were not considered further after correcting for multiple testing.

We also found few SNPs to be associated with more than one trait. The SNP S1_15880246 was significantly associated with iron (p-value 3.46E-06; FDR 0.002) and potassium (p-value 1.61E-04; FDR 0.08) content. Similarly, SNP S1_55137635 was associated with iron (p-value 4.02E-05; FDR 0.008) as well as zinc content (p-value 4.86E-06; FDR 0.03). Potassium and zinc content also shared a common association with S1_15078958 (p-value 4.47E-04; FDR 0.08 and p-value 2.97E-05; FDR 0.05, respectively). Among the 2085 MTAs identified through the GLM, about 330 reached the highly stringent genome-wide significance after Bonferroni correction for multiple testing (p < .05/73419 ≈ 6.81E-07/ −log10 p ≥ 6.17). These included 192 MTAs for iron, 70 for potassium, three for magnesium, 60 for sodium, four for zinc, and one for calcium content (Dataset S1). Similarly, from all the 430 associations detected using MLM approach, 34 MTAs (22 for iron, four for potassium, six for sodium and two for magnesium) also crossed the Bonferroni threshold (Dataset S1). Out of all 418 markers commonly

TABLE 4 Correlations between six grain micronutrients and protein measured in the association panel

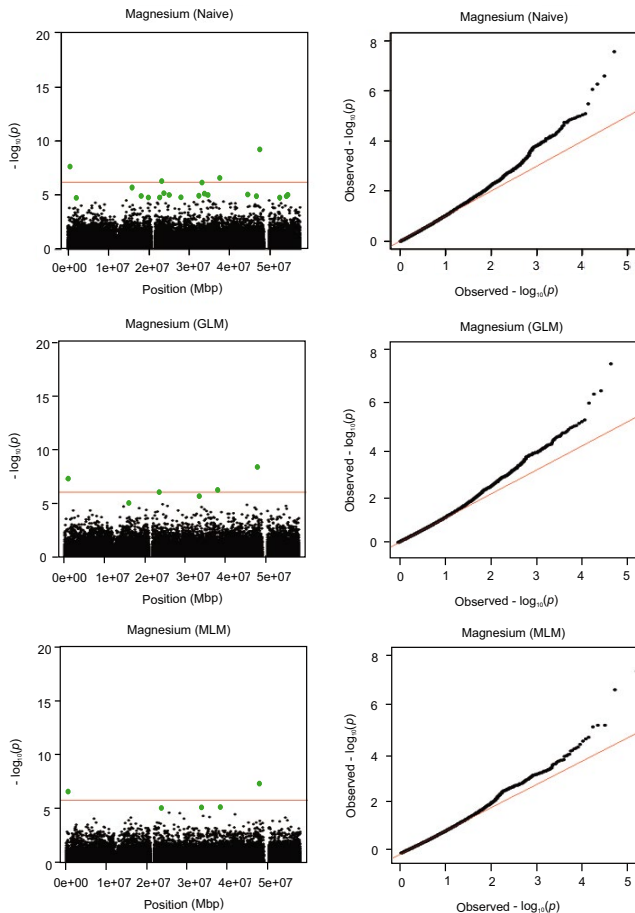|	Iron	Magnesium	Calcium	Potassium	Protein	Sodium	Zinc
Iron	1.00						
Magnesium	0.22[**]	1.00					
Calcium	0.16	0.41[*]	1.00				
Potassium	−0.05	−0.38[*]	−0.30[*]	1.00			
Protein	0.08	0.24[**]	−0.11	−0.05	1.00		
Sodium	0.06	−0.05	−0.06	0.28[*]	−0.10	1.00	
Zinc	0.10	0.27[*]	−0.03	−0.06	0.32[*]	−0.19[**]	1.00

*Significant at p < .001.

**Significant at p < .01.

FIGURE 4 Manhattan and corresponding quantile–quantile (QQ) plots of genome-wide association data showing – log10 *p*-value versus random position of SNP markers generated after relaxed filtering from the three models (a) Naïve (b) GLM (Q) (c) MLM (Q + K) for grain magnesium content. In the Manhattan plots, significant SNPs ($p \leq .001$, FDR < 0.1) are shown as circles in red, nonsignificant SNPs (FDR > 0.1) are shown as black points. The red horizontal line represents the Bonferroni threshold with an estimate of genome-wide *p*-value at 6.81E-07

genic sequences, whereas four belonged to chromosomal regions/scaffolds. Furthermore, from the seven SNPs associated with calcium, we found that three SNPs (S1_4620123, S1_44130155, and S1_5982733) encoded for orthologous cDNAs in other species. The majority of orthologues were observed in other monocot species such *Setaria italica*, *Oryza sativa*, *Zea mays*, *Ergostis tef*, *Brachypodium distachyon*, *Oropetium thomaeum*, *Sorghum bicolor*, *Panicum hallii*, and *Aegilops tauschii*. Many of these were predicted to have roles in metal ion binding, metal remobilization, or detoxification (Table S6).

4 | DISCUSSION

GBS has been used efficiently in millets with poorly assembled genomes, such as pearl millet, to generate a huge repository of SNPs for conducting various analysis such as genetic diversity and/or GWAS (Hu et al., 2015; Sehgal et al., 2012). Using this robust, multiplexed, high-throughput and low-cost GBS technique, 169,365 genome-wide high-quality SNPs were generated in our study. We therefore successfully showed that GBS can be used to generate a large number of high-quality markers in orphan species like finger millet, where marker number is currently limited.

The DAPC-based population structure analysis divided the population into three well-defined clusters related to their inherent genetic differences, which was mostly associated with their geographic origin. Irrespective of the germplasm collection, that is, among both the minicore/core and the elite germplasm, a clear distinction between the subpopulations enriched in accessions from Asia and those from parts of Africa was revealed. Such differentiation based on geographic pattern has also been previously reported (Kumar, Sharma, et al., 2016; Ramakrishnan, Ceasar, Duraipandiyan, Al-Dhabi, & Ignacimuthu, 2016). This division was further supported by Fst values. Pairwise comparisons among the two African subpopulations (cluster 1 and cluster 3) showed lower Fst values than those among the East African and Asian subpopulations (cluster 1 and cluster 2), or between South African and Asian subpopulations (cluster 3 and cluster 2). Thus, the less diverse genetic background of the East and South African accessions shows that they have a common evolutionary lineage and might have evolved from same natural population with its primary center of origin in Africa (Harlan & De Wet, 1971). Grouping of accessions from major finger millet growing regions (for example, in East African cluster countries such as Kenya, Uganda, and Tanzania) could be attributed to their close proximity geographically. This grouping is indicative of the conservation of a common gene pool between the primary and secondary center of origin of finger millet resulting in higher rates of gene flow in between the member countries (Hilu, de Wet, & Harlan, 1979). Unlike Bharathi's study (2011), however, European or American accessions were not grouped together into a single cluster. Marker density used in previous studies may have too poor to reveal these genetic groups. As Kumar, Sharma, et al. (2016) highlighted previously, some overlap of genotypes with accessions from other countries may be due to a big exchange and the ensuing hybridization and selection of Indian and African germplasm leading to allelic reshuffle among indigenous germplasm.

identified across GLM and MLM models, the 34 high confidence MTAs were used to perform homology search with other plants. In addition, as finger millet is known for high calcium content in its grains, we also performed homology search using the seven SNPs found to be associated with grain calcium content in the GLM.

Due to the current lack of a high-quality complete assembly of the finger millet whole genome sequence, it was difficult to directly estimate the genomic location of identified associations. Hence, based on an in silico comparative mapping approach using the 64 bp SNP harbouring sequences (Table S5), we performed database searches across various platforms such as Phytozome, NCBI, and Ensemble Plants server. From the 34 high confidence MTAs, 22 had orthologous sequences in other plants above the set E value ≤ 0.01 threshold (Table S6). The remaining 12 SNPs did not show any hits in the genomes of other monocots. From those having an orthologous region, 18 were those of a predicted mRNA or

The minicore and core germplasm sets used in this study have been well characterized previously and possess ample variability for several morphological and agronomic traits as well as grain nutrient content (Upadhyaya et al., 2006, 2011). We specifically included 48 locally adapted elite genotypes in the set to establish a direct comparison (of minerals and protein contents) with the genotypes of the core and minicore collections. In agreement with the prior studies, we found high variation for all seven traits in this study, suggesting that our association panel has sufficient variation to be effectively used for GWAS of various grain quality traits. However, with the exception of the sodium content of elite genotype KNE 622, most of the minicore/core genotypes surpassed the trait values of that of elite varieties. This is possibly the result of the modern elite varieties being developed through breeding programs aimed at improving specific traits, such as grain yield, rather than for grain nutritional content. As the traditional objective of agriculture systems and public policies has been to improve crop yields rather than their nutritional content, such high yielding cultivars often suffer a tradeoff between quality and quantity (Graham, Senadhira, Beebe, Iglesias, & Monasterio, 1999). Higher micronutrient concentration, however, does not always lead to grain yield penalty as evidenced by grain yield remaining unlinked to higher grain iron and zinc content in wheat (Welch & Graham, 2004) and pearl millet (Gupta, Velu, Rai, & Sumalini, 2009; Rai, Govindaraj, & Rao, 2012). Such results suggest that simultaneous selection for higher micronutrient contents without compromising on grain yield is possible. Genotypes of the finger millet minicore/core with better nutritional content can be used as donor genotypes in crosses for finger millet improvement.

A decisive factor for design of GWAS is the systematic characterization of the LD patterns in the genome (Serba et al.., 2019). LD is affected by several factors including fertilization behavior, rate of recombination, selection pressure, genetic drift, physical linkage, population structure, etc. Being a self-pollinated species, the lower rate of recombination in finger millet allows for relatively larger haplotype blocks. A previous study in finger millet showed that 17.9% of SNP marker pairs had significant LD at R^2 > 0.05 (Sharma et al., 2018). In other self-crossing millets such as foxtail millet, the genome-wide LD is reported to range from 100 Kb-177 Kb (Jaiswal et al., 2019; Jia et al., 2013). In the polyploid *Arabidopsis kamchatica*, the mean LD decay was found to be 5–10 kb, similar to *A. thaliana* and *Medicago truncatula* (LD decay within2-10 kb; Branca et al., 2011; Cao et al., 2011). The exact LD pattern of markers could not be identified here due to the lack of physical mapping distances between markers. The requirement for a more detailed analysis to assess LD and efficient QTL analysis may benefit from the recently published whole-genome draft sequence and assembly of finger millet (Hatakeyama et al., 2018).

QTL studies in finger millet have utilized association mapping based on low numbers of genic or genomic SSR markers. Such studies have allowed the identification of QTLs in finger millet including three for resistance to finger blast, three for leaf blast, one for neck blast (Babu, Dinesh, et al., 2014), five for agro-morphological characters (Babu, Agrawal, Pandey, Jaiswal, et al., 2014), seven for leaf blast resistance, tiller number, root length, seed yield (Ramakrishnan, Ceasar,

Duraipandiyan, Vinod, et al., 2016) and four for phosphorus response traits (Ramakrishnan et al., 2017). Principally, the selfing nature of finger millet, and the low rate of LD decay, should render a low marker density sufficient for identifying candidate genes within a larger genomic region (Ramakrishnan et al., 2017). Using this theory, a recent study found 109 novel SNPs to be associated with important agro-morphological traits such as grain yield in finger millet (Sharma et al., 2018). With only a few reports of genome-wide SNP markers and no reported MTA for micronutrient content in finger millet, we proceeded to generate a unique panel of SNPs and conduct a targeted GWAS. We attempted to identify MTAs between several new GBS-based SNP markers and a set of traits that are considerably crucial for human nutritional value. In total, 418 MTAs for four out of seven traits were identified. These SNPs can be considered robust as they were retained in both GLM and MLM analysis, although the latter model was more effective in controlling for confounding by population structure. Any spurious false positive associations due to multiple testing were controlled by applying FDR. The results also suggested that some of the significant MTAs were not detected by MLM because they did not reach the FDR criteria. It has often been reported that stringent FDR correction can sacrifice genuine MTAs as false negatives (Jaiswal et al., 2016, 2019; Kulwal et al., 2012). Thus, the MTAs identified for grain zinc and calcium content may still be worthy of further study and validation. Candidate genes for iron, sodium, potassium, and magnesium have not been previously reported in finger millet so our results present a novel finding. The MTAs identified through such genome level profiling are critical to initiate the identification of donor genotypes carrying desirable trait to act as divergent parents to be utilized in finger millet breeding. The recent release of the finger millet whole genome sequence will serve as a powerful reference tool in the coming years. As the rate of LD decay becomes known and the genomic location of clusters of significant SNPs can be revealed, the data presented in this study can be precisely used to conduct high-throughput GWAS. In terms of marker density, although the number of SNPs used in this study could not discover the entire range of QTLs, an improved GWAS resolution will also benefit from a higher marker density that covers nearly every haplotype block to properly map finger millet. For example, an increase in SNP density in rice genotyping arrays from 44,100 SNPs to 700,000 SNPs, and a further imputation-based augmentation to 5.2 million SNPs, has markedly improved the genotype-phenotype associations and functional polymorphisms underlying many of the QTLs identified by GWAS (McCouch et al., 2016; Wang et al., 2018; Zhao et al., 2011). Thus, the dissection of accurate QTLs and associated SNPs will prove highly beneficial for MAS of grain nutritional traits of finger millet.

Furthermore, we used the markers to detect candidate loci underlying the complex trait of grain nutrient content and accumulation. Due to the lack of whole-genome sequence information, however, the exact genomic position of the identified loci could not be predicted. Unlike, the candidate gene-based approach employed by Babu, Agrawal, Pandey, Jaiswal, et al. (2014), Babu, Dinesh, et al. (2014) and Nirgude et al. (2014), we utilized the comparative genomics approach to identify genomic regions that remain conserved across genomes. This method has also previously proved useful in finger

millet GWAS to delineate putative candidate genes for grain yield, flowering time and time to maturity (Sharma et al., 2018). The underlying sequences of 64bp harboring the SNP tag provided a relatively small sequence size to query against the NCBI/Phytozome/Ensembl databases. An orthologous predicted mRNA or genic sequence was present for 18 SNPs. In this respect, probably the most interesting candidate genes affecting these traits are those that include having predicted molecular function in metal ion binding or transport. One example, S1_30253617, which was found to be associated with iron content, was similar to a foxtail millet uncharacterized protein with a No apical meristem-associated (NAM) protein. NAM, a member of the NAC protein family (Puranik, Sahu, Srivastava, & Prasad, 2012), has been reported to play a role in iron and zinc remobilization to seeds during leaf senescence (Ricachenevsky, Menguer, & Sperotto, 2013). On studying iron content associated with marker S1_23343453, a probable mitochondrial 3-hydroxyisobutyrate dehydrogenase-like 1 (LOC101754224), homologous to that in *Setaria italica* was identified. The mammalian homologue of this gene is involved in accumulation and trafficking of intracellular iron (Devireddy, Hart, Goetz, & Green, 2010; Liu, Velpula, & Devireddy, 2014). The identification of associations with a number of markers indicate that there may be several genes functioning to remobilize, traffic, and maintain the levels of grain iron content in finger millet. We also found an the sodium associated marker, S1_53281655, to be a possible homologue of uncharacterized Et_s7379-1.39–1.mrna1 from a close relative of finger millet, *Ergostis tef*. This gene encoded a transcript containing Tetratricopeptide repeat, believed to have a role in providing salinity tolerance (Rosado et al., 2006). It is possible that this kind of gene may work to maintain intracellular osmotic ratio and protect the seed from ionic imbalance. Some transcripts were identified with functions involved in phenotypic response such as root development, regulation of plant inflorescence architecture or those involved in defense response to stress, flavonoid biosynthesis, proteolysis, and regulation of transcription. Interestingly, of the seven SNPs associated with calcium content, the SNP S1_5982733 encoded a SEUSS-like transcriptional corepressor. It is known to be involved in several developmental process which may often occur in a calcium-dependent manner, as known in mammals (Kashani et al., 2006). There were no previous studies that report the role of these genes in finger millet, hence it is imperative to validate these significant MTAs in future studies to ascertain the mechanism of transport, homeostasis, and allocation of grain minerals from the source to sink organs. Postvalidation, these SNPs can be utilized in full-length gene cloning and aid the introgression of favorable alleles into locally well-adapted germplasm through marker-assisted breeding.

5 | CONCLUSIONS

Our study highlights the potential of a large set of GBS-derived SNP markers for identifying population structure and GWA mapping of six grain MNs and protein content in finger millet for the first time. We included a diverse set of finger millet germplasm, including several African elite varieties which have never been utilized before in any such study. The study uncovers several novel MTAs and underlying candidate genes for grain mineral content in finger millet that until now remained unidentified. Without further knowledge of LD decay and reference genome the true genomic location of the identified QTLs cannot currently be reported. However, with the finger millet genome sequence information being recently available in the public domain, the associated markers identified in this study could be used in a much more precise way as validated markers might be used in finger millet breeding programs. The work provides an opportunity to use this sequence variation for identification and better characterization of novel alleles and genotypes and help to breed more nutritious finger millet genotypes. Promising alleles can provide great leads in this direction for future marker-assisted selection or for candidate gene cloning. Several of the elite varieties identified as having lower nutritional content may be taken up as target breeding material by exploiting the existing variability for further improving its value. The work can be a route to understand genetic pathways underlying high nutritional value of the finger millet grains. Moreover, from the human nutritional perspective, they can be encouraging candidates to improve the delivery the recommended daily intake of MNs benefiting the population as a whole and particularly currently marginalized communities.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHORS' CONTRIBUTIONS

RY and SP involved in design of the research. SP involved in performance of the research. SP, PPS, RKS, and HO carried out data collection and analysis. SP, PPS, SB, RKS, DS, and RY involved in data interpretation. SP, PPS, SB, DS, and RKS contributed to the writing of the manuscript.

ORCID

Swati Puranik https://orcid.org/0000-0003-1787-1550
Pranav P. Sahu https://orcid.org/0000-0003-4801-6602
Rakesh K. Srivastava https://orcid.org/0000-0001-6715-0386
Deepmala Sehgal https://orcid.org/0000-0002-4141-1784
Rattan Yadav https://orcid.org/0000-0001-6132-4541

REFERENCES

Agrama, H. A., Eizenga, G. C., & Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, *19*, 341–356. https://doi.org/10.1007/s11032-006-9066-6

Albert, E., Segura, V., Gricourt, J., Bonnefoi, J., Derivot, L., & Causse, M. (2016). Association mapping reveals the genetic architecture of tomato response to water deficit: Focus on major fruit quality traits. *Journal of Experimental Botany*, *67*, 6413–6430. https://doi.org/10.1093/jxb/erw411

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

AOAC. (2016). Latimer Jr. GW ed. *Official methods of analysis* (20th ed). Maryland, USA: AOAC International.

Babu, B. K., Agrawal, P. K., Pandey, D., Jaiswal, J. P., & Kumar, A. (2014). Association mapping of agro-morphological characters among the global collection of finger millet genotypes using genomic SSR markers. *Molecular Biology Reports*, *41*, 5287–5297. https://doi.org/10.1007/s11033-014-3400-6

Babu, B. K., Agrawal, P. K., Pandey, D., & Kumar, A. (2014). Comparative genomics and association mapping approaches for opaque2 modifier genes in finger millet accessions using genic, genomic and candidate gene-based simple sequence repeat markers. *Molecular Breeding*, *34*, 1261–1279.

Babu, B. K., Dinesh, P., Agrawal, P. K., Sood, S., Chandrashekara, C., Bhatt, J. C., & Kumar, A. (2014). Comparative genomics and association mapping approaches for blast resistant genes in finger millet using SSRs. *PLoS One*, *9*, e99182. https://doi.org/10.1371/journal.pone.0099182

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bharathi, A. (2011). *Phenotypic and genotypic diversity of global finger millet (Eleusine coracana (L.) Gaertn.) composite collection*. PhD thesis, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., ... Gentzbittel, L. et al (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences*, *108*, E864–E870.

Buckler, E. S., Holland, J. B., Bradbury, P. J., Acharya, C. B., Brown, P. J., Browne, C., ... McMullen, M. D. (2009). The genetic architecture of maize flowering time. *Science*, *325*, 714–718. https://doi.org/10.1126/science.1174276

Buckler, E. S., & Thornsberry, J. M. (2002). Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology*, *5*, 107–111. https://doi.org/10.1016/S1369-5266(02)00238-8

Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., ... Weigel, D. (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nature Genetics*, *43*, 956–963. https://doi.org/10.1038/ng.911

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Devireddy, L. R., Hart, D. O., Goetz, D. H., & Green, M. R. (2010). A mammalian siderophore synthesized by an enzyme with a bacterial homolog involved in enterobactin production. *Cell*, *141*, 1006–1017. https://doi.org/10.1016/j.cell.2010.04.040

Dhanapal, A. P., & Crisosto, C. H. (2013). Association genetics of chilling injury susceptibility in peach (*Prunus persica* (L.) Batsch) across multiple years. *3 Biotech*, *3*, 481–490. https://doi.org/10.1007/s13205-012-0109-x

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*, e19379.

Etter, P., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology (Clifton, NJ)*, *772*, 157–178.

FAO and International Life Science Institute (1997). *Preventing micronutrient malnutrition a guide to food-based approaches - Why policy makers should give priority to food-based strategies*. Washington, D.C. USA: ILSI PRESS. Retrieved from http://www.fao.org/3/x0245e/x0245e01.htm#P38_2721

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS One*, *9*, e90346. https://doi.org/10.1371/journal.pone.0090346

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., ... Rokhsar, D. S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Research*, *40*(D1), D1178–D1186. https://doi.org/10.1093/nar/gkr944

Goron, T. L., & Raizada, M. N. (2015). Genetic diversity and genomic resources available for the small millet crops to accelerate a new green revolution. *Frontiers in Plant Science*, *6*, 157. https://doi.org/10.3389/fpls.2015.00157

Graham, R. D., Senadhira, D., Beebe, S., Iglesias, C., & Monasterio, I. (1999). Breeding for micronutrient density in edible portions of staple food crops: Conventional approaches. *Field Crops Research*, *60*, 57–80. https://doi.org/10.1016/S0378-4290(98)00133-6

Gupta, S. M., Arora, S., Mirza, N., Pande, A., Lata, C., Puranik, S., ... Kumar, A. (2017). Finger Millet: A "certain" crop for an "uncertain" future and a solution to food insecurity and hidden hunger under stressful environs. *Frontiers in Plant Science*, *8*, 643. https://doi.org/10.3389/fpls.2017.00643

Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S., & Prasad, M. (2014). Population structure and association mapping of yield contributing agronomic traits in foxtail millet. *Plant Cell Reports*, *33*, 881–893. https://doi.org/10.1007/s00299-014-1564-0

Gupta, S. K., Velu, G., Rai, K. N., & Sumalini, K. (2009). Association of grain iron and zinc content with grain yield and other traits in pearl millet (*Pennisetum glaucum* (L.) R.Br.). *Crop Improvement*, *36*, 4–7.

Hao, D., Chao, M., Yin, Z., & Yu, D. (2012). Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (*Glycine max*) landraces. *Euphytica*, *186*, 919–931. https://doi.org/10.1007/s10681-012-0697-x

Harlan, J. R., & de Wet, J. M. J. (1971). Towards a rational classification of cultivated plants. *Taxon*, *20*, 509–517.

Hatakeyama, M., Aluri, S., Balachadran, M. T., Sivarajan, S. R., Patrignani, A., Grüter, S., ... Francoijs, K. J. et al (2018). Multiple hybrid de novo genome assembly of finger millet, an orphan allotetraploid crop. *DNA Research*, *25*, 39–47.

Hilu, K. W., de Wet, J. M. J., & Harlan, J. R. (1979). Archeobotany and the origin of finger millet. *American Journal of Botany*, *66*, 330–333.

Hittalmani, S., Mahesh, H. B., Shirke, M. D., Biradar, H., Udat, G., Aruna, Y. R., ... Mohanrao, A. (2017). Genome and transcriptome sequence of finger millet (*Eleusine coracana* (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. *BMC Genomics*, *18*, 465. https://doi.org/10.1186/s12864-017-3850-z

Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., ... Flicek, P. (2020). Ensembl Genomes 2020-enabling

non-vertebrate genomic research. *Nucleic Acids Research*, *48*, D689–D695. https://doi.org/10.1093/nar/gkz890

Hu, Z., Mbacké, B., Perumal, R., Guèye, M. C., Sy, O., Bouchet, S., ... Morris, G. P. (2015). Population genomics of pearl millet (*Pennisetum glaucum* (L.) R. Br.): Comparative analysis of global accessions and Senegalese landraces. *BMC Genomics*, *16*, 1048. https://doi.org/10.1186/s12864-015-2255-0

Jaiswal, V., Gahlaut, V., Meher, P. K., Mir, R. R., Jaiswal, J. P., Rao, A. R., ... Gupta, P. K. (2016). Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*T. aestivum* L.). *PLoS One*, *11*, e0159343.

Jaiswal, V., Gupta, S., Gahlaut, V., Muthamilarasan, M., Bandyopadhyay, T., Ramchiary, N., & Prasad, M. (2019). Genome-wide association study of major agronomic traits in foxtail millet (*Setaria italica* L.) using ddRAD sequencing. *Scientific Reporter*, *9*, 5020. https://doi.org/10.1038/s41598-019-41602-6

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., ... Han, B. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nature Genetics*, *45*, 957–961. https://doi.org/10.1038/ng.2673

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, *11*, 94. https://doi.org/10.1186/1471-2156-11-94

Kashani, A. H., Qiu, Z., Jurata, L., Lee, S. K., Pfaff, S., Goebbels, S., ... Ghosh, A. (2006). Calcium activation of the LMO4 transcription complex and its role in the patterning of thalamocortical connections. *Journal of Neuroscience*, *26*, 8398–8408. https://doi.org/10.1523/JNEUROSCI.0618-06.2006

Kooke, R., Kruijer, W., Bours, R., Becker, F., Kuhn, A., van de Geest, H., ... Keurentjes, J. J. B. (2016). Genome-wide association mapping and genomic prediction elucidate the genetic architecture of morphological traits in Arabidopsis. *Plant Physiology*, *170*, 2187–2203. https://doi.org/10.1104/pp.15.00997

Kulwal, P. L. (2016). Association mapping and genomic selection—where does sorghum stand? In: S. Rakshit, & Y. H. Wang, eds. *The sorghum genome: Compendium of plant genomes*. Cham, Switzerland:Springer.

Kulwal, P., Ishikawa, G., Benscher, D., Feng, Z., Yu, L.-X., Jadhav, A., ... Sorrells, M. E. (2012). Association mapping for pre-harvest sprouting resistance in white winter wheat. *Theoretical and Applied Genetics*, *125*, 793–805. https://doi.org/10.1007/s00122-012-1872-0

Kumar, A., Metwal, M., Kaur, S., Gupta, A. K., Puranik, S., Singh, S., ... Sood, S. et al (2016). Nutraceutical value of finger millet [*Eleusine coracana* (L.) Gaertn.], and their improvement using omics approaches. *Frontiers. Plant Science*, *7*, 934.

Kumar, A., Sharma, D., Tiwari, A., Jaiswal, J. P., Singh, N. K., & Sood, S. (2016). Genotyping-by-sequencing analysis for determining population structure of finger millet germplasm of diverse origins. *The Plant Genome*, *9*, 1–15. https://doi.org/10.3835/plantgenome2015.07.0058

Li, D., Zhao, X., Han, Y., Li, W., & Xie, F. (2018). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics*, pii: S0888-7543(18), 30008–30009.

Liu, Z., Velpula, K. K., & Devireddy, L. (2014). Hydroxybutyrate dehydrogenase-2 and ferritin-H synergistically regulate intracellular iron. *FEBS Journal*, *281*, 2410–2421.

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., ... Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, *9*, e1003215.

Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and jones' factors. *Critical Reviews in Food Science and Nutrition.*, *48*, 177–184. https://doi.org/10.1080/10408390701279749

McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., ... Mezey, J. (2016). Open access resources for genome-wide association mapping in rice. *Nature Communications*, *7*, 10532. https://doi.org/10.1038/ncomms10532

McLean, E., Cogswell, M., Egli, I., Wojdyla, D., & De Benoist, B. (2009). Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993–2005. *Public Health Nutrition*, *12*, 444–454.

Nirgude, M., Babu, B. K., Shambhavi, Y., Singh, U. M., Upadhyaya, H. D., & Kumar, A. (2014). Development and molecular characterization of genic molecular markers for grain protein and calcium content in finger millet (*Eleusine coracana* (L.) Gaertn.). *Molecular Biology Reports*, *41*, 1189–1200. https://doi.org/10.1007/s11033-013-2825-7

Padulosi, S., Mal, B., Ravi, S. B., Gowda, J., Gowda, K. T. K., Shanthakumar, G., ... Dutta, M. (2009). Food security and climate change: Role of plant genetic resources of minor millets. *Indian Journal of Plant Genetic Resources*, *22*, 1–16.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Pasam, R. K., Sharma, R., Malosetti, M., van Eeuwijk, F. A., Haseneyer, G., Kilian, B., & Graner, A. (2012). Genome-wide association studies for agronomical traits in a worldwide spring barley collection. *BMC Plant Biology*, *12*, 16. https://doi.org/10.1186/1471-2229-12-16

Puranik, S., Kam, J., Sahu, P. P., Yadav, R., Srivastava, R. K., Ojulong, H., & Yadav, R. (2017). Harnessing finger millet to combat calcium deficiency in humans: Challenges and prospects. *Frontiers in Plant Science*, *8*, 1311. https://doi.org/10.3389/fpls.2017.01311

Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). NAC proteins: Regulation and role in stress tolerance. *Trends in Plant Science*, *17*, 1360–1385. https://doi.org/10.1016/j.tplants.2012.02.004

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*, 559–575. https://doi.org/10.1086/519795

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rai, K. N., Govindaraj, M., & Rao, A. S. (2012). Genetic enhancement of grain iron and zinc content in pearl millet. *Quality Assurance and Safety of Crops and Foods*, *4*, 119–125. https://doi.org/10.1111/j.1757-837X.2012.00135.x

Ramakrishnan, M., Ceasar, S. A., Duraipandiyan, V., Al-Dhabi, N. A., & Ignacimuthu, S. (2016). Assessment of genetic diversity, population structure and relationships in Indian and non-Indian genotypes of finger millet (*Eleusine coracana* (L.) Gaertn) using genomic SSR markers. *SpringerPlus*, *5*, 120.

Ramakrishnan, M., Ceasar, S. A., Duraipandiyan, V., Vinod, K. K., Kalpana, K., Al-Dhabi, N. A., & Ignacimuthu, S. (2016). Tracing QTLs for leaf blast resistance and agronomic performance of finger millet (*Eleusine coracana* (L.) Gaertn.) genotypes through association mapping and *in silico* comparative genomics analyses. *PLoS One*, *11*, e0159264. https://doi.org/10.1371/journal.pone.0159264

Ramakrishnan, M., Ceasar, S. A., Vinod, K. K., Duraipandiyan, V., Ajeesh Krishna, T. P., Upadhyaya, H. D., ... Ignacimuthu, S. (2017). Identification of putative QTLs for seedling stage phosphorus starvation response in finger millet (*Eleusine coracana* L. Gaertn.) by association mapping and cross species synteny analysis. *PLoS One*, *12*, e0183261. https://doi.org/10.1371/journal.pone.0183261

Ricachenevsky, F. K., Menguer, P. K., & Sperotto, R. A. (2013). kNACking on heaven's door: How important are NAC transcription factors for leaf senescence and Fe/Zn remobilization to seeds? *Frontiers in Plant Science*, 4, 226. https://doi.org/10.3389/fpls.2013.00226

Rosado, A., Schapire, A. L., Bressan, R. A., Harfouche, A. L., Hasegawa, P. M., Valpuesta, V., & Botella, M. A. (2006). The Arabidopsis Tetratricopeptide Repeat-Containing Protein TTL1 is required for osmotic stress responses and abscisic acid sensitivity. *Plant Physiology*, 142, 1113–1126. https://doi.org/10.1104/pp.106.085191

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4, 406–425.

Sehgal, D., Rajaram, V., Armstead, I. P., Vadez, V., Yadav, Y. P., Hash, C. T., & Yadav, R. S. (2012). Integration of gene-based markers in a pearl millet genetic map for identification of candidate genes underlying drought tolerance quantitative trait loci. *BMC Plant Biology*, 12, 9. https://doi.org/10.1186/1471-2229-12-9

Sehgal, D., Skot, L., Singh, R., Srivastava, R. K., Das, S. P., Taunk, J., ... Yadav, R. S. (2015). Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. *PLoS One*, 10, e0122165. https://doi.org/10.1371/journal.pone.0122165

Serba, D. D., Muleta, K., St. Amand, P. C., Bernardo, A., Bai, G., Perumal, R., & Bashir, E. (2019). Genetic diversity, population structure, and linkage disequilibrium of pearl millet. *Plant Genome*, 12, 180091.

Sharma, D., Tiwari, A., Sood, S., Jamra, G., Singh, N. K., Meher, P. K., & Kumar, A. (2018). Genome wide association mapping of agro-morphological traits among a diverse collection of finger millet (*Eleusine coracana* L.) genotypes using SNP markers. *PLoS One*, 13, e0199444. https://doi.org/10.1371/journal.pone.0199444

Shashi, B. K., Sharan, S., Hittalamani, S., Shankar, A. G., & Nagarathna, T. K. (2007). Micronutrient composition, antinutritional factors and bioaccessibility of iron in different finger millet (*Eleusine coracana*) genotypes. *Karnataka Journal of Agricultural Sciences*, 20, 583–585.

Sood, S., Kumar, A., Babu, B. K., Gaur, V. S., Pandey, D., Kant, L., & Pattnayak, A. (2016). gene discovery and advances in finger millet [*Eleusine coracana* (L.) Gaertn.] genomics—an important nutri-cereal of future. *Frontiers Plant Science*, 7, 1634. https://doi.org/10.3389/fpls.2016.01634

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences USA*, 100, 9440–9445. https://doi.org/10.1073/pnas.1530509100

Tripathi, B., & Platel, K. (2010). Finger millet (*Eleucine coracana*) flour as a vehicle for fortification with zinc. *Journal of Trace Elements in Medicine and Biology*, 24, 46–51. https://doi.org/10.1016/j.jtemb.2009.09.001

Turner, S. D. (2018). qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3, 731. https://doi.org/10.21105/joss.00731

Upadhyaya, H. D., Gowda, C. L. L., Pundir, R. P. S., Reddy, V. G., & Singh, S. (2006). Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits. *Genetic Resources and Crop Evolution*, 53, 679–685. https://doi.org/10.1007/s10722-004-3228-3

Upadhyaya, H. D., Gowda, C. L. L., & Reddy, V. G. (2007). Morphological diversity in finger millet germplasm introduced from Southern and Eastern Africa J. *Journal of SAT Agricultural Research*, 3, 1–3.

Upadhyaya, H. D., Ramesh, S., Sharma, S., Singh, S. K., Varshney, S. K., Sarma, N., ... Singh, S. (2011). Genetic diversity for grain nutrients contents in a core collection of finger millet (*Eleusine coracana* (L.) Gaertn.) germplasm. *Field Crops Research*, 121, 42–52. https://doi.org/10.1016/j.fcr.2010.11.017

Upadhyaya, H. D., Sarma, N. D. R. K., Ravishankar, C. R., Albrecht, T., Narasimhudu, Y., Singh, S. K., ... Gowda, C. L. L. (2010). Developing a mini-core collection in finger millet using multilocation data. *Crop Science*, 50, 1924–1931. https://doi.org/10.2135/cropsci2009.11.0689

Vetriventhan, M., Upadhyaya, H. D., Dwivedi, S. L., Pattanashetti, S. K., & Singh, S. K. (2015). Finger and foxtail millets. In M. Singh, & H. D. Upadhyaya (Eds.), *Genetic and genomic resources for cereals improvement*. London, UK: Academic press

Wang, D. R., Agosto-Pérez, F. J., Chebotarov, D., Shi, Y., Marchini, J., Fitzgerald, M., ... McCouch, S. R. (2018). An imputation platform to enhance integration of rice genetic resources. *Nature Communications*, 9, 3519. https://doi.org/10.1038/s41467-018-05538-1

Welch, R. M., & Graham, R. D. (2004). Breeding for micronutrients in staple food crops from a human nutrition perspective. *Journal of Experimental Botany*, 55, 353–364. https://doi.org/10.1093/jxb/erh064

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., ... Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42, 565–569. https://doi.org/10.1038/ng.608

Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17, 155–160. https://doi.org/10.1016/j.copbio.2006.02.003

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., ... Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203–208. https://doi.org/10.1038/ng1702

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., ... Mezey, J. et al (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature Communications*, 2, 467.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.