

Whole-Genome Sequencing and Analysis of Nutritional Traits in a Core Set of Ethiopian Teff (*Eragrostis tef*) Varieties

Nelzo C Ereful

NIAB

Huw Jones

NIAB

Nick Fradgley

NIAB

Lesley Boyd

NIAB

Hirut Assaye Cherie

Bahir Dar Institute of Technology

Matthew J Milner (✉ matthew.milner@niab.com)

NIAB

Research Article

Keywords:

Posted Date: January 27th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1251010/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Teff (*Eragrostis teff*) is a tropical cereal domesticated and grown in the Ethiopian highlands, where it has been a staple food of Ethiopians for many centuries. Food insecurity and nutrient deficiencies are major problems in the country, so breeding for enhanced nutritional traits, such as Zn content, could help to alleviate problems with malnutrition.

Results: To understand the breeding potential of nutritional traits in teff a core set of 24 varieties were sequenced and their mineral content, levels of phytate and protein, as well as a number of nutritionally valuable phenolic compounds measured in grain. Significant variation in all these traits was found between varieties. Genome wide sequencing of the 24 teff varieties revealed 3,193,582 unique SNPs and 897,272 unique INDELS relative to the teff reference var. Dabbi. Sequence analysis of two key transporter families involved in the uptake and transport of Zn by the plant led to the identification of 32 Zinc Iron Permease (ZIP) transporters and 14 Heavy Metal Associated (HMA) transporters in teff. Further analysis identified numerous variants, of which 14.6% of EtZIP and 12.4% of EtHMA variants were non-synonymous changes. Analysis of a key enzyme in flavanol synthesis, flavonoid 3'-hydroxylase (F3'H), identified a T-G variant in the teff homologue Et_s3159-0.29-1.mrna1 that was associated with the differences observed in kaempferol glycoside and quercetin glycoside levels.

Conclusion: Wide genetic and phenotypic variation was found in 24 Ethiopian teff varieties which would allow for breeding gains in many nutritional traits of importance to human health.

Background

Teff (*Eragrostis teff*) is a tropical cereal that has its origins in the Ethiopian highlands, where it was domesticated and has been grown for thousands of years [1]. Globally teff is a minor cereal crop in terms of both production and planted area, with Ethiopia growing an estimated 90% of the annual global teff crop on about three million hectares, which equates to about a quarter of the Ethiopian grain-cultivated area [2]. Compared to other cereals teff is considered a resilient crop. It can withstand adverse weather conditions, growing well at elevations between 1,800 and 2,200m above sea level, in regions where there is adequate rainfall [3]. Teff is a staple of Ethiopian diets, providing 11% of the per capita caloric intake and two-thirds of the average Ethiopians daily protein [4].

Teff is widely considered a healthy alternative to cereals such as wheat, maize and rice as it does not contain gluten, is high in slowly digestible starch, rich in calcium (Ca) and polyphenols [5,6]. However, levels of zinc (Zn) in the grain (in the range of 28-40 mg/kg) are often less than the recommended level of 40 mg/kg necessary to meet human nutritional requirements [1,7,8], with further fortification efforts being hampered by low Zn levels in many Ethiopian soils [1].

While breeding efforts to improve teff have been ongoing in Ethiopia since the mid-1950s germplasm advancements have been slow, with only 20 new teff varieties being released [3]. A better understanding of teff nutritional diversity and its genetic control could help drive nutritional gains for this crop. Recent

advancements in our understanding of the genetics of teff include a fairly complete genome assembly, the release of gene models, as well as a number of RNASeq datasets which are all publicly available [9]. These allow for translation of knowledge on grain fortification from other species, supporting more rapid advancements in efforts to improve the nutritional potential of teff [10–14].

Several gene families have been identified that support accumulation and transport of various heavy metals within the plant. The Zinc Iron Permease (ZIP) family of transporters, the Heavy Metal Associated (HMA) and Metal Tolerance Protein (MTP) families, and the Natural Resistance-Associated Macrophage protein (NRAMPs) family of transporters have all been shown to be integral in micronutrient transport in plants [15–19]. The transport of Zn from soil to seed can involve a number of these transporters, with Zn primarily being accumulated within the seed embryo and aleurone layer [20]. In rice Zn uptake from the soil is thought to occur through the transporters *OsZIP5* and *OsZIP9* [21,22]. Others have suggested that many routes may be available to Zn uptake, as no single transporter is believed to be solely responsible for the uptake of Zn from the soil [23–27]. Several transporter families have also been implicated in the transport of Zn through the plant to the developing grains [16,28,29]. These include the HMA family and the ZIP family of transporters, both shown to be able to transport a broad range of metals [16,17,24,30], which can also include the transport of unwanted metals such as cadmium (Cd) and lead (Pb), both of which are detrimental to most plants and animals [15,17,24,30–32].

Other factors can also influence the Zn content in both plant leaves and grain. These include phytates, which readily bind cations such as calcium (Ca), iron (Fe) and Zn, phosphate (P) and phytate being found to correlate with Zn concentration in the grain of both wheat and rice [29,33]. In wheat, nitrogen (N) content has been found to positively correlate with Zn content in the grain [8]. Recent work has shown that macronutrients such as N, P and sulphur (S), can also influence overall micronutrient content [34]. Thus, a holistic approach to nutrient enhancement is required to ensure adequate Zn supply in the human diet.

In this study we set out to understand the phenotypic variation of several nutritional traits in teff varieties commonly grown and consumed in Ethiopia. We used whole genome sequencing and single nucleotide polymorphisms (SNPs) to determine the genetic relationship between these teff varieties. This genomic data was then used to look for relationships between the teff varieties and the nutritional traits to identify those varieties with optimal nutritional potential for breeding. In addition, SNPs within candidate genes underlying nutritional traits of interest were examined to determine linkages with the trait variation.

Materials And Methods

Plant material and flour preparation

Twenty-four teff varieties (Additional file 1) were grown in the main cropping season in 2018 to 2019 at the experimental station of Adet Agricultural Research Center, Ethiopia. Adet (11°28' N, 37°48' E; 2216 m a.s.l) is located 42 km southwest of Bahir Dar, the capital city of the Amhara regional state, Ethiopia. Agronomic and cultural practices recommended for teff production were which included application of 40

kg N and 60 kg P₂O₅ per ha [35]. The soil at Adet is brown Nitosol. The lines which have been used in the study are publicly available lines in Ethiopia from the Adet Agricultural Research Center. The voucher specimen ID numbers are listed in additional file 1.

Teff whole grains were milled to flour using a Laboratory Hammer Mill (Model: TPS-JXFM110, China) to a sieve size of 0.5 mm, packed in polyethylene bags and stored at room temperature until analysed.

Analyses of nutritional traits

Analyses of all nutritional traits were undertaken at NIAB, UK. Analyses included assessment of elemental concentrations, phenolic compounds, phytate and N content.

Elemental concentrations were assessed using inductive coupled plasma mass spectrometry (ICP-MS). ICP-MS was used to detect the beneficial micronutrients Ca, Cu, Fe, K, Mg, Mn, Mo, P, S, Se, Ti, and Zn, and the detrimental elements Cd and Pb. Approximately 0.2–0.3 g of flour was digested in 5 mL of nitric acid (Sigma) overnight in a 7 mL vial. The digested sample was transferred to a 50 mL beaker and heated to 115°C to remove the residual acid, after which 3 mL of H₂O₂ (Sigma) was added. After H₂O₂ reduction the remaining powder was dissolved in 15 mL of ddH₂O and samples analysed on the ICP-MS using a Thermo-Fisher Scientific iCAP-Q equipped with CCTED (collision cell technology with energy discrimination). Three, independent technical replications were run for each teff flour sample. Possible soil contamination was identified where flour samples had greater than 100 mg/kg of Fe, Al and possibly Si, and/ or Ti above 1 µg/g [36,37].

N content was determined using Dumas analysis. Flour samples were dried at 104°C for three hours. One gram of flour was loaded according to the manufacturer's instructions (Leco TruMacN Dumas gas analyser) to determine N content. Dumas gas analysis was performed on three, technical replicate 1 g aliquots of flour from each variety.

Phenolic compounds were assessed using HPLC. Flour samples of approx. 2.5 g were extracted into 50 ml of ethanol-acetic acid (10% 1M acetic acid v/v) under reflux conditions for 2 hours. Extracts were stored at -20 °C until analysed. The extracts were prepared for chromatography by centrifugation for two minutes at 13000 rpm, then filtered through a 0.2 µm filter. The compounds were separated using the Dionex Ultimate 3000 HPLC system. A 150 mm x 4.6 mm x 5 µm x 100 Å Kinetix C18 column was used, with a gradient 0.1% formic acid / acetonitrile mobile phase running at 0.2 ml/minute (gradient of 0.95: 0.05 for 2 minutes, then 0.72; 0.28 for 18 minutes, 0.00; 0.10 for 28 minutes, and then held until 45 minutes) were used. The column effluent was monitored with a PDA detector between 200 - 600 nm, with data recorded at 254 nm, 280 nm, 340 nm and 520 nm.

Phytate was measured using the commercial Megazyme Phytic Acid Assay Kit (Brey, Ireland) following the manufacturer's instructions with minor modifications. Approximately 100 mg of flour was digested in 1.8 mL HCl (0.66 M) in 2.2 mL tubes, placed in a rotator mixer overnight with a constant rpm of 20, at room temperature. Three technical replicates were applied for each flour sample.

Genomic DNA isolation and sequencing

DNA was extracted from flour of the 24 teff varieties using Qiagen's DNAeasy Kit as per the manufacturer's instructions, including RNase treatment. The same flour samples were used for DNA extraction as were used for nutrient trait analyses. DNA was shipped to Novogene (Cambridge UK) for Illumina Sequencing. Illumina libraries contained on average 350 bp inserts and were sequenced using paired end technology. The estimated genome coverage of each teff variety was at least 25X, based on an estimated size of the genome of 622 MB [9]. Raw reads for all teff varieties tested have been deposited with the ENA under the ArrayExpress accession E-MTAB-8827.

Whole genome sequence assembly and variant calling

Paired-end sequence reads were provided by the sequencing service that had already undergone quality checks, including the removal of over-represented sequences, adapters, and reads with low-quality base scores (at $Q > 20$). No further analysis using FastQC [38] was therefore required. The paired-end reads were mapped against the indexed reference sequence of the teff variety Dabbi using BWA mem [9,39]. Output was piped to bam and sorted using SAMtools. PICARD tool was used to assign reads to a single Read Group (option: AddOrReplaceReadGroup). PCR duplicates were flagged using PICARD (option: MarkDuplicates). A dictionary for the reference genome of contig names and sizes was also created using PICARD (option: CreateSequenceDictionary). Variants were identified using the Broad Institute Genome Analysis Tool Kit (GATK4; [40]). GATK (option: Haplotypecaller) was run for each teff variety, using its default parameters, to call variants between the reference Dabbi genome sequence and the variety sequence in both Variant Call Format (VCF) and Genomic VCF (option: -ERC GVCF) modes. Only variants with a quality score of > 20 were considered. The variant SNPs were then analysed using gif3 and bcftools csq [41] to determine whether the SNP was within a coding region, and whether the SNP was synonymous vs non-synonymous.

Using PLINK2 the VCF file was converted into PLINK binary format (bed, bim and fam). A phylogenetic tree was created using SNPhylo and the biallelic SNPs identified across the 24 teff varieties [42]. A principal co-ordinate analysis (PCA) plot was constructed using the same biallelic SNPs and analysed using the R program SNPrelate [43]. FastStructure (with logistic prior, $K = 5$) was used to infer population structure within the 24 teff varieties using the same SNPs [44].

Identification of zinc iron permease and heavy metal associated transporter family members in teff

The *Eragrostis tef* genome was searched using the Biomart tool in Ensembl plants (<http://plants.ensembl.org/biomart/martview/>) to identify predicted genes that contain both the PFAM domain PF02535 and Interpro ID IPR003689, or Interpro ID IPR027256. The PFAM domain PF02535 and Interpro ID IPR003689 are used for the identification of ZIP transporter proteins and Interpro ID IPR027256 for identification of HMA transporter proteins. Candidate ZIP transporter and HMA transporter proteins in *Arabidopsis* and rice were also identified using Ensembl's Biomart. The amino acid sequences of these candidate genes were aligned using MUSCLE as part of MEGA X [45] using the default

parameters and a phylogenetic tree created using the Maximum Likelihood option, with 50 replications used for the determination of bootstrap values.

Statistical analysis

Micronutrient, N and phytate levels were tested for normality and homogeneity of variance using Shapiro-Wilk's test in R, which indicated a normal distribution. Two-way analysis of variance (ANOVA) was used to infer significance between teff varieties using the aov function and TukeyHSD commands in R. A least significant difference of 5% probability level was used as a post-hoc test to determine significance. Results were plotted using R ggplot2 and ggpubr packages [46–48]. Correlations between micronutrients, N and phytate were calculated in R using the Spearman rank correlation (cor.test) or Kruskal-Wallis test for categorical variables [43].

All statistical analyses of phenolic compound levels were performed using Genstat v.16 (VSN International 2020). The levels of the phenolic compounds were analysed using a modified two-way ANOVA approach, General Linear Regression. The model applied was replicate by variety. Only comparisons having a F probability < 0.001 were considered as statistically significant. The linear relationship between phenolic compound levels were measured using the Pearson correlation coefficient in Excel (Microsoft). Boxplots were generated using R ggplot2 [47].

Results

Comparison of nutritional traits in flour of twenty-four teff varieties

To understand the nutritional potential of 24 teff varieties (Additional file 1) currently grown and consumed in Ethiopia we measured the levels of a number of nutritional traits, comparing the relationship between these traits, and using genomic data to look at the genetic relationship between these 24 teff varieties relative to these nutritional traits. The nutritional traits included elemental micronutrient (Fig. 1 & Additional file 2), nitrogen as a proxy for protein content (Fig. 2), phytate (Fig. 2) and a range of phenolic compounds (Additional file 3).

Significant variation between the 24 teff varieties was found for all micronutrients tested (Additional file 2) (p -val < 0.001). Focus on elements essential for human health, Zn and Fe (Figure 1), showed Zn flour concentrations to range from 14.8 (var Etsub) to 29.2 mg/kg (var. Heber-1) and Fe to range from 22.6 (var Abay) to 684.25 mg/kg (var Yilmana) (Figure 1A & B). Positive correlations were found between Zn and Ca, Mg, P, S, and to a lesser extent K, while negative correlations were found with Cu, Fe, Mo and Se. A positive correlation was also seen between Zn and Cd (Suppl. Figure 1). No correlations were found between Pb or Ti and the other elements measured.

Cd is a toxic micronutrient, detrimental to human health. High levels of Cd were seen in many of the teff samples tested, the highest levels of Cd being found in the variety Wellenkomi, having 40.7 µg/kg.

Only one of the teff varieties, var Magna, had Cd levels below the current EU limits for Cd in food products, being below 1 µg/kg (1ppm)[49].

The N levels within the teff flours were measured as a proxy for protein content[50]. Significant differences were found between the teff varieties, with N values ranging from 1.3 g/100g in var. Ambo Toketo nearly 2 g/100g in var. Dagan teff ($p < 0.001$) (Figure 2A). Using the standard conversion of 5.95 this would give protein content in the range of 7.79% to 11.71%[51]. A positive correlation was found between Zn content and N levels in the flours tested ($R = 0.59$; $p < 0.01$; Figure 2B) but not for Fe and N ($R = 0.0096$; $p < 0.97$; Figure 2C).

Significant differences were found in phytate levels between the 24 teff varieties ($p < 0.01$) (Figure 2D). The phytate levels ranged from 0.83 g/100g in var. Magna to 2.56 g/100g in var. Abay. No significant correlations were observed between phytate and Zn ($R = 0.042$; $p < 0.84$; Figure 2E), but a significant negative correlation was found between Fe and phytate ($R = -0.54$; $p < 0.007$; Figure 2F). There was no significant correlation between overall P levels, measured via ICP-MS, and phytate ($R = 0.29$; $p < 0.18$).

Flour of the teff varieties were also screened for 19 phenolic compounds (Additional file 3). Significant differences ($F < 0.001$) were found between the 24 teff varieties for 11 phenolic compounds, significant differences between teff varieties not being found for gallic acid. Seven phenolic compounds were not detected. These include cyanidin, cyanidin glycoside, delphinidin, delphinidin glycoside, flavone, pelargonidin and p-coumaric acid.

Kaempferol, quercetin, catechin and myricetin all belong to a group of phenolic compounds commonly known as flavonoids (Figure 3)[52]. The levels of kaempferol and quercetin were low in all the teff varieties, the majority of these compounds being present as kaempferol glycoside and quercetin glycoside (Figure 4). The levels of kaempferol glycoside and quercetin glycoside ranged from 4.92 (var. Tseday) to 205.79 µg/g (var. Magna), and 0.46 (var. Felagot) to 105.41 µg/g (var. Quncho), respectively. The levels of catechin ranged from 27.41 (var. Yilmana) to 183.19 µg/g (var. Tseday) (Figure 4). Catechin is derived from dihydroxyquercetin (Figure 3) and a precursor of proanthocyanidins, which are thought to give rice its red colouration [53]. The levels of myricetin ranged from 0.82 (var. Were-kiyu) to 17.24 µg/g (var. Dima) (Figure 4).

The levels of *t*-cinnamic acid ranged from 1.51 µg/g in var. Baset to 24.65 µg/g in var. Dukem. Cinnamic acid is a precursor of ferulic acid, which in turn gives rise to vanillic acid. A strong positive correlation was found between cinnamic and ferulic acids, and a lesser positive correlation with vanillic acid. However, the levels of vanillic acid were low in all 24 teff varieties (Additional file 3, Additional file 4).

Protocatechuic acid and gallic acid are synthesised from a side branch of the shikimate pathway that leads to the synthesis of folates and aromatic amino acids, including phenylalanine (Figure 3; [53]). The levels of protocatechuic acid ranged from 6.78 µg/g (var. Areka-1) to 44.78 µg/g (var. Felagot) (Additional file 3). The levels of gallic acid were low in all 24 teff varieties, ranging from 0.12 µg/g (var. Dima) to 0.97 µg/g (var. Ambo take) (Figure 4).

Pearson's correlation analyses were undertaken on the levels of each phenolic compound, phytate and Fe (Additional file 4). A strong, positive correlation was seen between cinnamic acid and ferulic acid ($r = 0.78$), fitting with the biochemical pathway (Figure 3), where cinnamic acid is a precursor of ferulic acid. Similarly, a positive association was seen between ferulic acid and vanillic acid ($r = 0.51$), vanillic acid sitting downstream of ferulic acid. However, positive correlations were also seen between ferulic acid and myricetin ($r = 0.61$), which may relate to their common precursor cinnamic acid, and between cinnamic acid and quercetin ($r = 0.49$) and quercetin glycoside ($r = 0.50$).

The precursors of kaempferol and quercetin, dihydrokaempferol and dihydroquercetin, exist in an equilibrium controlled by a flavonoid 3'-hydroxylase (F3'H) enzyme which converts dihydrokaempferol to dihydroquercetin. This resulted in a negative correlations between kaempferol and quercetin ($r = -0.28$) and kaempferol glycoside and quercetin glycoside ($r = -0.43$); the majority of kaempferol and quercetin being present in the glycosylated state. This was reflected in positive correlations between kaempferol and quercetin glycoside ($r = 0.62$) and quercetin and kaempferol glycoside ($r = 0.45$).

Catechin is derived from dihydroquercetin and therefore competes for synthesis with quercetin. Consequently, a negative correlation was observed between the levels of catechin and quercetin ($r = -0.32$). However, a positive correlation was seen between catechin and quercetin glycoside ($r = 0.52$). A positive correlation was also observed between catechin and kaempferol ($r = 0.57$) and a negative correlation with kaempferol glycoside ($r = -0.51$).

In addition to positive correlations with kaempferol ($r = 0.62$) and catechin ($r = 0.52$), positive correlations were observed between quercetin glycoside and other phenolic compounds, including ferulic acid ($r = 0.65$), myricetin ($r = 0.56$), cinnamic acid ($r = 0.50$). These correlations may indicate a positive feedback mechanism operating through the biosynthetic pathways leading to quercetin glycoside synthesis.

A negative correlation was found between gallic acid and salicylic acid ($r = -0.43$), and between catechin and protocatechuic acid ($r = -0.44$), while a positive correlation was observed between protocatechuic acid and Fe content ($r = 0.40$). No significant correlations were found between phytate levels and any of the phenolic compounds measured in this study (Additional file 4).

Genetic relationship between the twenty-four teff varieties

All 24 teff varieties were sequenced to greater than 25X coverage. Sequences were compared to the reference var. Dabbi across the whole genome. The varieties differed by as few as 1.567 million SNPs in var. Areka-1 relative to Dabbi, to as high as 2.372 million SNPs in var. Yilmana (Additional file 5). There was also considerable variation in INDELs, with Yilmana showing the fewest INDELs at 343,257 and Wellenkomi the highest at 520,234. Overall, 3,193,582 unique SNPs and 897,272 unique INDELs were found, containing a minimum allele frequency of 0.1 within the 24 varieties tested. Considering that the whole teff genome is estimated to be 622 Mb in size, the 2.372 million variants identified in var. Yilmana relative to the reference var. Dabbi, equate to roughly four variants in every kb of the genome.

To understand how the 24 teff varieties relate to each other a phylogenetic tree analysis (Figure 5A), a principal component analysis (PCA; Figure 5B), and a structure analysis (Figure 5C) were carried out using the genome wide SNP data. The phylogenetic tree and PCA separated the 24 teff varieties into 3 groups, however the structure analysis returned two groupings. In general, the teff varieties fell into similar groupings when comparing the phylogenetic tree and PCA, the exception being the varieties Negus and Kora, which fell on the same branch of the tree but in distinct PCA groups.

Genetic variation underlying differences in Zn levels

Zn concentrations were overlaid on the PCA to see if a relationship between flour Zn levels and the teff variety groupings could be identified (Figure 5B). As no specific association between genetic grouping of teff varieties and Zn concentration were apparent, we chose to look at the genetic variation in specific gene families involved in Zn transport. Two gene families were selected. The ZIP (Zinc Iron Permease) family are involved in uptake of Zn from the soil and the HMA (Heavy metal associated) family of transporters are involved in movement of Zn from roots to seeds.

Zinc Iron Permease transporter family in teff

To identify putative ZIP family members the gene models of the *Eragrostis tef* var. Dabbi sequence in Ensembl plants was searched using the PFAM domain PF02535 and Interpro ID IPR003689. This revealed 32 predicted genes in teff which contained these protein domains. As teff is a tetraploid species these 32 potential ZIP family members is comparable to the 15 in *Arabidopsis* and 17 in rice [30,54]. However, some of the predicted genes could be pseudogenes. Only six of the 32 coding regions identified have a good ATG start codon, and two of these six putative ZIP transporters lacked a stop codon. However, as this incomplete sequence data could be due to gaps in the reference var. Dabbi teff sequence we used all 32 putative ZIP sequences in subsequent analyses.

Phylogenetic analysis of the translated, amino acid sequences of the 32 teff ZIP transporters was performed with members of the ZIP families from rice and *Arabidopsis*. (Figure 6). Teff ZIP proteins showed closer linkages with rice ZIP proteins compared to *Arabidopsis*, both teff and rice being monocots. In most cases there were two teff genes for every rice ZIP gene, which is expected as teff is a tetraploid species, but not all 32 teff ZIP genes showed clear associations with genes found in rice or *Arabidopsis*, suggesting some divergence. Rice ZIP genes associated with more than two teff genes included OsZIP8, which was associated with four teff genes. Only one of the 32 teff ZIP genes localized with OsZIP3, while three teff genes colocalized with OsIRT2, as well as OsZIP10.

Heavy Metal Associated transporter family in teff

To identify putative HMA family members the gene models of the *Eragrostis tef* var. Dabbi sequence in Ensembl plants was searched using the Interpro domain IPR027256. Fourteen HMA proteins were identified in teff compared to eight in *Arabidopsis* and nine in rice. Phylogenetic analysis of the translated, amino acid sequences of the 14 teff HMA transporters was performed with members of the

HMA families from rice and *Arabidopsis* (Figure 7). As 14 teff genes is less than expected for this tetraploid species, based on conservation of each of the eight core genes found in both rice and *Arabidopsis*, the teff Dabbi reference sequence may be incomplete with regards to this family of Zn transporters. In addition, five of 14 predicted HMA genes in teff did not contain an ATG start codon, suggesting that some sequence is missing, or the gene models are incorrect. Four rice HMA proteins, OsHMA2, OsHMA5, OsHMA6 and OsHMA9, have two clear homologues with predicted teff HMA genes, but there does not appear to be any identified teff HMA gene aligning with OsHMA3, a major gene involved in Cd tolerance and sequestration [55].

Identification of variants in ZIP and HMA proteins

Numerous SNPs were found in both the ZIP and HMA teff families of transporters (Additional files 6, Additional file 7). In the 32 *EtZIP* genes a total of 355 variants were identified in the coding regions relative to the reference var. Dabbi, this included frame shift mutations in twelve of the 32 genes (3.4% of the total mutations) (Additional file 6). Most of the variants were located in an intron (41.4%), followed by synonymous variants in the coding region (22.5%), and then by non-synonymous mutations that result in a change in the amino acid sequence (14.6%). Many of the variants were conserved between the varieties, including a frameshift in both homologues of OsZIP9, with most of the varieties containing the mutated/truncated form of each gene.

In *EtHMA* transporters 298 variants were identified relative to var. Dabbi (Additional file 7). Most of the variants were again found in introns or represented synonymous mutations in the coding sequence (48% and 16.4%, respectively). Only two of the 14 teff HMA genes had frame shift variants. This included the predicted gene loci Et_s3091-2.42-1.mrna1 and Et_s3193-0.29-1.mrna1. The third most common variants were non-synonymous mutations in the protein coding region (12.4%).

Genetic variation underlying differences in phenolic compounds

Flavonoid 3'-hydroxylase (F3'H) is a key enzyme in the conversion of dihydrokaempferol to dihydroquercetin [53,56]. To determine whether F3'H in teff was responsible for the differences seen between teff varieties in kaempferol glycoside and quercetin glycoside levels the amino acid sequence of the rice F3'H gene (Os10g0320100) was used to identify possible orthologs in teff. Using an e-value cut off of 1e-10 five putative orthologs of OsF3'H were identified in the teff reference sequence of var. Dabbi: Et_s9738-1.8-1.mrna1, Et_s9399-0.5-1.mrna1, Et_s3159-0.29-1.mrna1, Et_s6352-0.10-1.mrna1 and Et_s15942-0.0-1.mrna1 (Suppl. Figure 2). Alignment of the amino acid sequences of these five teff genes to their rice putative orthologs showed 58.74% to 78.57% identity. The two teff genes Et_s3159-0.29-1.mrna1 and Et_s15942-0.0-1.mrna1 clustered most closely with Os10g0320100 and were subject to further analysis. SNPs were identified between Et_s3159-0.29-1.mrna1 and Et_s15942-0.0-1.mrna1. Correlations between these SNPs and the levels of kaempferol glycoside and quercetin glycoside in the 24 teff varieties revealed a SNP in Et_s3159-0.29-1.mrna1 that was strongly correlated with kaempferol glycoside and quercetin glycoside levels (p val < 0.001). This SNP resulted in a T G substitution in the second intron of Et_s3159-0.29-1.mrna1, and therefore does not directly change the coding sequence of

the gene (Suppl. Figure 2). Teff varieties containing the wild-type (WT) “T” allele in the homozygous state had lower levels of kaempferol glycoside than teff varieties containing the mutant “G” SNP, and vice versa, while varieties with a heterozygous SNP had intermediary levels of the two glycosides (Figure 8). No further associations were found between the other phenolic compounds measured and SNPs in any of the F3’H-type genes in teff.

Discussion

Teff is a staple crop for many Ethiopians, however, due to naturally low Zn levels in soils on which it is grown the grain does not meet the Zn needs of those who consume it [57]. Zn is essential for a healthy immune system and **stimulates the activity of many different enzymes**. The low level of Zn in teff has therefore led to many Ethiopian children suffering from Zn deficiency [58]. Fe is also an essential element, required for normal blood cell function [59], but has a low bioavailability, the small intestine not readily absorbing large amounts of Fe. However, a number of compounds have been shown to influence Fe absorption in the human gut, including phytate, kaempferol and quercetin [60–62].

To understand the breeding potential of teff for Zn and Fe content, as well as several other nutritional factors, grain from 24 teff varieties were assayed for a range of nutritional traits. Analysis of these 24 varieties showed significant variation in all the micronutrients tested, with a two-fold difference in the amount of Zn being observed. However, even var. Heber-1, with 29.2 mg/kg of Zn, had levels below the recommended 40 mg/kg of Zn in flour [8]. This may be rectified by Zn fertilization of the soil, but future work would be required to understand how each variety responds to Zn soil supplements, given the complexity of Zn uptake and transportation to the grain.

No correlation was found between Zn and Fe levels in the grains, and is most likely due to different regulatory mechanisms and transporters involved in the movement of each ion through the plant [16,28]. The lack of identification of a homologue of *OsHMA3*, a gene associated with Cd tolerance and sequestration in rice, may also suggest an underlying cause for some of the high levels of Cd seen in the 24 teff flours. However, it should also be noted that other putative teff genes, including a homolog of *OsZIP1* (Et_s3548-1.55-1.mrna1) which is involved in efflux of excess heavy metals, was one of the twelve teff genes with a frame shift variant present in all the teff varieties. This would suggest that multiple genes which sequester Cd in the roots, and keep it away from the grain, may not exist in teff or are non-functional in the modern teff varieties assayed in this study [63].

Sequencing of the 24 teff varieties enabled the identification of several variants which could be used to develop markers for future breeding. Overall, there was a large amount of genetic variation between the varieties compared to the reference var. Dabbi. Although most of these variants were in non-coding regions or synonymous mutations, as seen in the analysis of the ZIP and HMA transporter families. The variation observed is not unexpected as similar levels of variation have been observed in a core set of rice varieties which were recently sequenced [64].

In rice, double mutants of *OsZIP5* and *OsZIP9* have shown severe Zn deficiency symptoms, suggesting these two genes are the major route for Zn into the rice plant [21,22]. However, mutants of either gene in isolation does not show major effects on Zn uptake. Mutations in two homologues of *OsZIP9*, including frame shifts, as well as a frame shift variant in a homologue of *OsZIP5* in some varieties, would suggest that the route of Zn into teff, while compromised by these mutations, is not wholly dependent on the homologues of *OsZIP5* and *OsZIP9*. Other transporters may be the major route for Zn into teff plants. These other routes might also contribute to the high levels of Cd seen in many of the teff varieties. It has been shown that genes involved in Mn and Fe uptake can also transport Cd in vivo whereas high affinity Zn transporters do not appear to have this capability [15,27,65–67].

Comparison of other nutritional traits, including nitrogen (as a proxy for protein), phytate and a number of phenolic compounds, also showed significant variation between the 24 teff varieties. The variation in the phenolic profiles is of importance as some phenolics have been found to alter Fe bioavailability and may therefore present a more effective strategy than Fe fortification for enhancing nutritional outcomes. It is often reported that phytate inhibits Fe bioavailability, so the negative correlation between Fe and phytate levels in teff could prove beneficial. While future research is required to determine whether elevated levels of phenolics that stimulate and inhibit Fe uptake in the human gut can help alleviate the anaemia seen in Ethiopian children [69], the significance of the SNP in the teff F3'H gene Et_s3159-0.29-1.mrna1, which explains a large proportion of the variation in kaempferol glycoside and quercetin glycoside levels, suggest this might be a breeding target to improve Fe bioavailability in teff, as kaempferol glycoside is known to promote Fe absorption while quercetin glycoside inhibits Fe absorption in cell assays [60]. In addition, phenolic compounds have also been found to influence other agronomic traits, including tolerance to both biotic and abiotic stress [14,60,61,68].

With the considerable variation seen for the nutritional traits assessed in this study breeders are in a good position to breed for enhanced nutritional value in teff. The genomic sequence information collected can be used to identify and develop markers linked to target genes and traits. We have yet to test the heritability of these traits and their stability over growing locations and seasons, but marker-assisted selection, particularly within target genes, can now provide a feasible approach to breed for these nutritional traits in teff.

Conclusion

For many subsistence Ethiopian farming families teff is a major crop and source of calories. Yet, as shown in this study, levels of Zn are usually below recommended levels, while levels of Cd exceed EU limits. However, considerable phenotypic and genetic variation for a range of nutritional traits and the genes regulating their levels in planta, exists. This provides considerable potential to determine the relationship between these nutrition phenotypes and identify allelic variants that would allow breeding of new teff varieties with optimal nutritional potential.

Abbreviations

N: Nitrogen

Zn: Zinc

P: Phosphate

Fe: Iron

Cd: Cadmium

Pb: Lead

Ca: Calcium

S: Sulphur

K: Potassium

Mg: Magnesium

Mn: Manganese

Mo: Molybdenum

Se: Selenium

Ti: Titanium

Al: Aluminium

Si: Silicon

HPLC: High-performance liquid chromatography

PDA: Photo Diode Array

HCl: Hydrochloric acid

Vcf: Variant Call Format

EU: European Union

MTP: Metal Tolerance Protein

ZIP: Zinc Iron Permease

HMA: Heavy Metal Associated

NRAMP: Natural Resistance-Associated Macrophage Proteins

SNP: Single Nucleotide Polymorphism

INDEL: Insertion or Deletion

ICP-MS: Inductively Coupled Plasma Mass Spectrometry

PCA: Principal Component Analysis

ENA: European Nucleotide Archive

Declarations

Seeds collected:

Permission to collect and analyse the milled seeds/flour of the teff varieties documented in this work was obtained from the Ethiopian government prior to research being conducted and samples being sent.

Ethics approval and consent to participate:

All the methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication:

All authors consent to publication

Availability of data and material:

Sequencing Data is available at the ENA under ArrayExpress accession E-MTAB-8827.

Competing interests:

The authors declare no competing interests.

Funding:

This work was supported by the Biotechnology and Biological Sciences Council (UK) BBSRC (grant BB/P027458/1) and pump priming grant (GCRF-SA-2020-NIAB). As well as the Cambridge Global Challenges fund. The funding bodies listed here played no role in the design of the study and collection, analysis, and interpretation of data, or in the writing of the manuscript.

Authors' contributions:

MM, NF, NE, and HAC designed the experiments. MM, NF, NE, HJ performed the experiments. NE undertook the bioinformatic analyses. MM and LB undertook the statistical analyses of the data. The manuscript

was written by MM and LB and subsequently revised by all authors. All authors have read and approved the final manuscript.

Acknowledgements:

The authors would like to thank the Adet Agricultural Research Center for providing the teff samples. The University of Nottingham school of Chemistry for running the ICP-MS samples for elemental content. Helen Appleyard and Marianna Rizzo at NIAB analytical services for analysis for N content analysis of the grains.

References

1. Baye K. Teff: nutrient composition and health benefits - IFPRI Publications - IFPRI Knowledge Collections [Internet]. Int. Food Policy Res. Inst. 2014 [cited 2020 Oct 27]. p. 20. Available from: <https://ebrary.ifpri.org/digital/collection/p15738coll2/id/128334/>
2. Lee H. Teff, A Rising Global Crop: Current Status of Teff Production and Value Chain. Open Agric J. Bentham Science Publishers Ltd.; 2018;12:185–93.
3. Fufa B, Behute B, Simons R, Berhe T. Strengthening the Tef Value Chain in Ethiopia. 2011.
4. Foodgrain consumption and calorie intake patterns in Ethiopia - IFPRI Publications - IFPRI Knowledge Collections [Internet]. [cited 2020 Oct 22]. Available from: <https://ebrary.ifpri.org/digital/collection/p15738coll2/id/124853>
5. Zhu F. Chemical composition and food uses of teff (*Eragrostis tef*). Food Chem. Elsevier Ltd; 2018;239:402–15.
6. Cheng A, Mayes S, Dalle G, Demissew S, Massawe F. Diversifying crops for food and nutrition security - a case of teff. Biol Rev Camb Philos Soc. Biol Rev Camb Philos Soc; 2017;92:188–98.
7. Abebe Y, Bogale, Alemtsehay Hambidge, K. Michael Stoecker BJ, Bailey K, Rosalind GS. Phytate, zinc, iron and calcium content of selected raw and prepared foods consumed in rural Sidama, Southern Ethiopia, and implications for bioavailability. J Food Compos Anal. 2007;20:161–8.
8. Cakmak I, Pfeiffer WH, McClafferty B. Biofortification of durum wheat with zinc and iron. Cereal Chem. John Wiley & Sons, Ltd; 2010. p. 10–20.
9. VanBuren R, Man Wai C, Wang X, Pardo J, Yocca AE, Wang H, et al. Exceptional subgenome stability and functional divergence in the allotetraploid Ethiopian cereal teff. Nat Commun. Nature Research; 2020;11:1–11.
10. Nakandalage N, Nicolas M, Norton RM, Hirotsu N, Milham PJ, Seneweera S. Improving Rice Zinc Biofortification Success Rates Through Genetic and Crop Management Approaches in a Changing Environment. Front Plant Sci. 2016;7.
11. Klein MA, Grusak MA. Identification of nutrient and physical seed trait QTL in the model legume *Lotus japonicus*. Genome. 2009;52:677–91.

12. Arora S, Cheema J, Poland J, Uauy C, Chhuneja P. Genome-wide association mapping of grain micronutrients concentration in *aegilops tauschii*. *Front Plant Sci. Frontiers Media S.A.*; 2019;10.
13. Oury FX, Leenhardt F, Rémésy C, Chanliaud E, Duperrier B, Balfourier F, et al. Genetic variability and stability of grain magnesium, zinc and iron concentrations in bread wheat. *Eur J Agron.* 2006;25:177–85.
14. Wright TIC, Gardner KA, Glahn RP, Milner MJ. Genetic control of iron bioavailability is independent from iron concentration in a diverse winter wheat mapping population. *BMC Plant Biol. BioMed Central Ltd*; 2021;21:1–13.
15. Milner MJ, Mitani-Ueno N, Yamaji N, Yokosho K, Craft E, Fei Z, et al. Root and shoot transcriptome analysis of two ecotypes of *Noccaea caerulescens* uncovers the role of *NcNramp1* in Cd hyperaccumulation. *Plant J.* 2014;78.
16. Milner M, Kochian L. Investigating Heavy-metal Hyperaccumulation using *Thlaspi caerulescens* as a Model System | *Annals of Botany | Oxford Academic.* *Ann Bot.* 2008;102:3–13.
17. Hussain D, Haydon MJ, Wang Y, Wong E, Sherson SM, Young J, et al. P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in arabidopsis. *Plant Cell. American Society of Plant Biologists*; 2004;16:1327–39.
18. Desbrosses-Fonrouge AG, Voigt K, Schröder A, Arrivault S, Thomine S, Krämer U. Arabidopsis thaliana MTP1 is a Zn transporter in the vacuolar membrane which mediates Zn detoxification and drives leaf Zn accumulation. *FEBS Lett. No longer published by Elsevier*; 2005;579:4165–74.
19. Dräger DB, Desbrosses-Fonrouge AG, Krach C, Chardonnens AN, Meyer RC, Saumitou-Laprade P, et al. Two genes encoding Arabidopsis halleri MTP1 metal transport proteins co-segregate with zinc tolerance and account for high MTP1 transcript levels. *Plant J. John Wiley & Sons, Ltd*; 2004;39:425–39.
20. De Brier N, Gomand S V, Donner E, Paterson D, Smolders E, Delcour JA, et al. Element distribution and iron speciation in mature wheat grains (*Triticum aestivum* L.) using synchrotron X-ray fluorescence microscopy mapping and X-ray absorption near-edge structure (XANES) imaging. *Plant Cell Environ.* 2016;39:1835–47.
21. Yang M, Li Y, Liu Z, Tian J, Liang L, Qiu Y, et al. A high activity zinc transporter *OsZIP9* mediates zinc uptake in rice. *Plant J. Wiley*; 2020;103:1695–709.
22. Tan L, Qu M, Zhu Y, Peng C, Wang J, Gao D, et al. Zinc transporter5 and zinc transporter9 function synergistically in zinc/cadmium uptake1. *Plant Physiol. American Society of Plant Biologists*; 2020;183:1235–49.
23. Milner MJ, Craft E, Yamaji N, Koyama E, Ma JF, Kochian L V. Characterization of the high affinity Zn transporter from *Noccaea caerulescens*, *NcZNT1*, and dissection of its promoter for its role in Zn uptake and hyperaccumulation. *New Phytol.* 2012;
24. Grotz N, Fox T, Connolly E, Park W, Guerinot M Lou, Eide D. Identification of a family of zinc transporter genes from Arabidopsis that respond to zinc deficiency. *Proc Natl Acad Sci U S A. National Academy of Sciences*; 1998;95:7220–4.

25. Tiong J, McDonald G, Genc Y, Shirley N, Langridge P, Huang CY. Increased expression of six *ZIP* family genes by zinc (Zn) deficiency is associated with enhanced uptake and root-to-shoot translocation of Zn in barley (*Hordeum vulgare*). *New Phytol.* Blackwell Publishing Ltd; 2015;207:1097–109.
26. Palmgren MG, Clemens S, Williams LE, Krämer U, Borg S, Schjørring JK, et al. Zinc biofortification of cereals: problems and solutions. *Trends Plant Sci.* Trends Plant Sci; 2008. p. 464–73.
27. Curie C, Alonso JM, Le Jean M, Ecker JR, Briat JF. Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport. *Biochem J.* Portland Press Ltd; 2000;347:749–55.
28. Sperotto RA, Ricachenevsky FK, Williams LE, Vasconcelos MW, Menguer PK. From soil to seed: Micronutrient movement into and within the plant. *Front. Plant Sci.* Frontiers Research Foundation; 2014.
29. Magallanes-López AM, Hernandez-Espinosa N, Velu G, Posadas-Romano G, Ordoñez-Villegas VMG, Crossa J, et al. Variability in iron, zinc and phytic acid content in a worldwide collection of commercial durum wheat cultivars and the effect of reduced irrigation on these traits. *Food Chem.* Elsevier Ltd; 2017;237:499–505.
30. Milner MJ, Seamon J, Craft E, Kochian L V. Transport properties of members of the ZIP family in plants and their role in Zn and Mn homeostasis. *J Exp Bot.* 2012;
31. Papoyan A, Piñeros M, Kochian L V. Plant Cd²⁺ and Zn²⁺ status effects on root and shoot heavy metal accumulation in *Thlaspi caerulescens*. *New Phytol.* John Wiley & Sons, Ltd; 2007;175:51–8.
32. Pence NS, Larsen PB, Ebbs SD, Letham DLD, Lasat MM, Garvin DF, et al. The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. *Proc Natl Acad Sci U S A.* National Academy of Sciences; 2000;97:4956–60.
33. Stangoulis JCRR, Huynh B-LL, Welch RM, Choi E-YY, Graham RD. Quantitative trait loci for phytate in rice grain and their relationship with grain micronutrient content. *Euphytica* Springer; p. 289–94.
34. Kumar S, Kumar S, Mohapatra T. Interaction Between Macro- and Micro-Nutrients in Plants. *Front Plant Sci.* Frontiers; 2021;0:753.
35. Abewa A, Adgo E, Yitaferu B, Alemayehu G, Assefa K, Solomon JKQ, et al. Teff Grain Physical and Chemical Quality Responses to Soil Physicochemical Properties and the Environment. *Agron* 2019, Vol 9, Page 283. Multidisciplinary Digital Publishing Institute; 2019;9:283.
36. Cherney JH, Robinson DL. A Comparison of Plant Digestion Methods for Identifying Soil Contamination of Plant Tissue by TI Analysis¹. *Agron J.* John Wiley & Sons, Ltd; 1983;75:145–7.
37. Jones JB. Plant Tissue Analysis in Micronutrients. *Micronutr Agric.* John Wiley & Sons, Ltd; 2018;477–521.
38. Andrews S. FastQC: a quality control tool for high throughput sequence data [Internet]. 2010. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
39. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.

40. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* Cold Spring Harbor Laboratory Press; 2010;20:1297–303.
41. Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Birol I, editor. *Bioinformatics.* Oxford University Press; 2017;33:2037–9.
42. Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics.* BioMed Central; 2014;15:162.
43. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* Oxford Academic; 2012;28:3326–8.
44. Raj A, Stephens M, Pritchard JK. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics.* Genetics; 2014;197:573–89.
45. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.
46. Kassambara A. ggplot2. Based Publ Ready Plots [R Packag ggpubr version 02 5]. 2020;
47. Hadley W. Ggplot2: Elegrant graphics for data analysis. Springer; 2016.
48. Computing RF for S. R Core Team (2018). R: A language and environment for statistical computing [Internet]. 2018. Available from: <https://www.r-project.org/>
49. EUR-Lex - 02006R1881-20150521 - EN - EUR-Lex [Internet]. [cited 2020 Sep 12]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02006R1881-20150521>
50. Heathcote JG. The Protein Quality of Oats. *Br J Nutr.* Cambridge University Press; 1950;4:145–54.
51. Fujihara S, Sasaki H, Aoyagi Y, Sugahara T. Nitrogen-to-protein conversion factors for some cereal products in Japan. *J Food Sci.* J Food Sci; 2008;73.
52. Mbanjo EGN, Kretzschmar T, Jones H, Ereful N, Blanchard C, Boyd LA, et al. The Genetic Basis and Nutritional Benefits of Pigmented Rice Grain. *Front. Genet.* Frontiers Media S.A.; 2020. p. 229.
53. Mbanjo EGN, Kretzschmar T, Jones H, Ereful N, Blanchard C, Boyd LA, et al. The Genetic Basis and Nutritional Benefits of Pigmented Rice Grain. *Front Genet.* Frontiers Media S.A.; 2020;11:229.
54. Chen WR, Feng Y, Chao YE. Genomic analysis and expression pattern of OsZIP1, OsZIP3, and OsZIP4 in two rice (*Oryza sativa* L.) genotypes with different zinc efficiency. *Russ J Plant Physiol.* 2008;55:400–9.
55. Sasaki A, Yamaji N, Ma JF. Overexpression of OsHMA3 enhances Cd tolerance and expression of Zn transporter genes in rice. *J Exp Bot.* Oxford University Press; 2014;65:6013–21.
56. Lam PY, Lui ACW, Yamamura M, Wang L, Takeda Y, Suzuki S, et al. Recruitment of specific flavonoid B-ring hydroxylases for two independent biosynthesis pathways of flavone-derived metabolites in grasses. *New Phytol.* Blackwell Publishing Ltd; 2019;223:204–19.
57. Tessema M, De Groote H, D. Brouwer I, J.M. Feskens E, Belachew T, Zerfu D, et al. Soil Zinc Is Associated with Serum Zinc But Not with Linear Growth of Children in Ethiopia. *Nutrients.*

- 2019;11:221.
58. Ayana G, Moges T, Samuel A, Asefa T, Eshetu S, Kebede A. Dietary zinc intake and its determinants among Ethiopian children 6–35 months of age. *BMC Nutr. BioMed Central*; 2018;4:30.
 59. Sciences NA of S (US) and NRC (US) D of M. THE ROLE OF IRON IN HEMOGLOBIN SYNTHESIS. National Academies Press (US); 1958;
 60. Hart JJ, Tako E, Glahn RP. Characterization of Polyphenol Effects on Inhibition and Promotion of Iron Uptake by Caco-2 Cells. *J Agric Food Chem. American Chemical Society*; 2017;65:3285–94.
 61. Hart JJ, Tako E, Kochian L V., Glahn RP. Identification of Black Bean (*Phaseolus vulgaris* L.) Polyphenols That Inhibit and Promote Iron Uptake by Caco-2 Cells. *J Agric Food Chem.* 2015;63:5950–6.
 62. Jin F, Frohman C, Thannhauser TW, Welch RM, Glahn RP. Effects of ascorbic acid, phytic acid and tannic acid on iron bioavailability from reconstituted ferritin measured by an in vitro digestion-Caco-2 cell model. *Br J Nutr. Br J Nutr*; 2009;101:972–81.
 63. Liu XS, Feng SJ, Zhang BQ, Wang MQ, Cao HW, Rono JK, et al. OsZIP1 functions as a metal efflux transporter limiting excess zinc, copper and cadmium accumulation in rice. *BMC Plant Biol. BioMed Central Ltd.*; 2019;19:1–16.
 64. Tanaka N, Shenton M, Kawahara Y, Kumagai M, Sakai H, Kanamori H, et al. Whole-Genome Sequencing of the NARO World Rice Core Collection (WRC) as the Basis for Diversity and Association Studies. *Plant Cell Physiol. Oxford University Press*; 2020;61:922–32.
 65. Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot M Lou, Briat JF, et al. IRT1, an Arabidopsis Transporter Essential for Iron Uptake from the Soil and for Plant Growth. *Plant Cell. Oxford University Press*; 2002;14:1223.
 66. Lin YF, Hassan Z, Talukdar S, Schat H, Aarts MGM. Expression of the Znt1 zinc transporter from the metal hyperaccumulator *Noccaea caerulescens* confers enhanced zinc and cadmium tolerance and accumulation to *Arabidopsis thaliana*. *PLoS One. Public Library of Science*; 2016;11.
 67. Milner MJ, Craft E, Yamaji N, Koyama E, Ma JF, Kochian L V. Characterization of the high affinity Zn transporter from *Noccaea caerulescens*, NcZNT1, and dissection of its promoter for its role in Zn uptake and hyperaccumulation. *New Phytol. New Phytol*; 2012;195:113–23.
 68. Singh P, Arif Y, Bajguz A, Hayat S. The role of quercetin in plants. *Plant Physiol Biochem. Elsevier Masson*; 2021;166:10–9.
 69. Hailu A. Ethiopian National Micronutrient Survey Report: September 2016. *Ethiop Public Heal Institute, Minist Heal.* 2016;

Figures

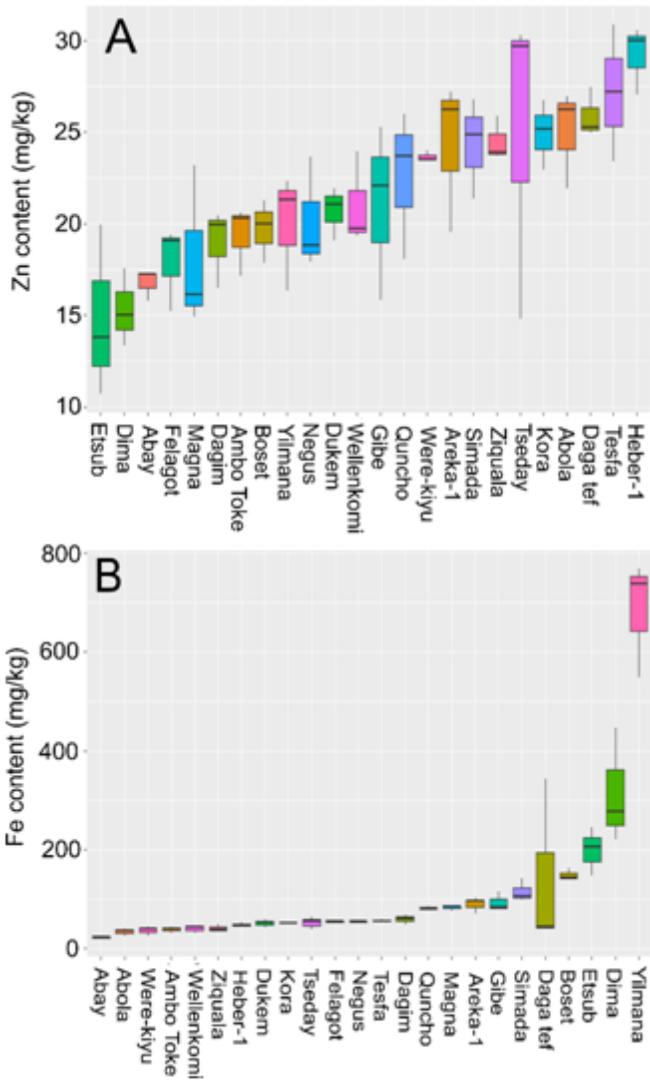


Figure 1

Zn and Fe levels in flour of 24 teff varieties grown at the Adet Research Agricultural Research Center in 2018/19. The box defines the upper and lower quartiles. The lines extending vertically, “whiskers”, indicate variability outside the upper and lower quartiles.

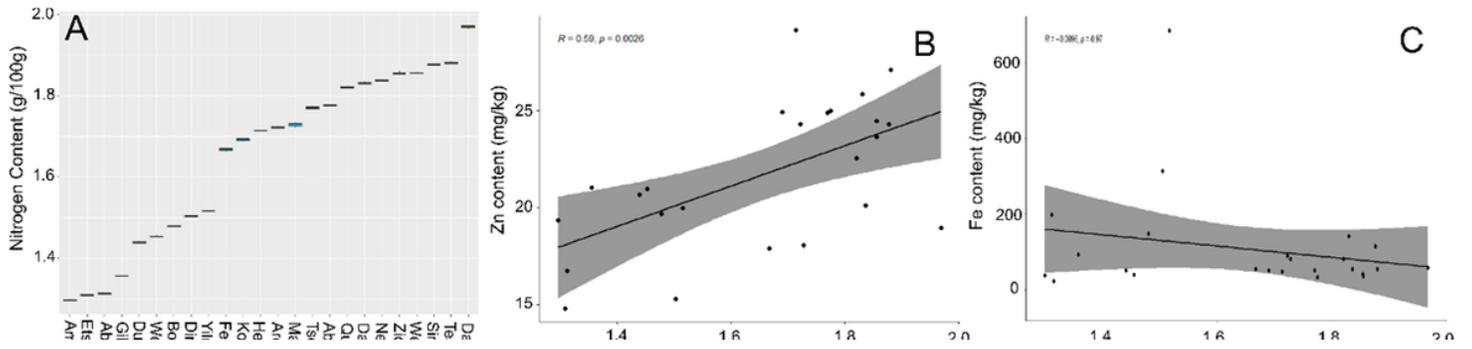


Figure 2

Relationship between Zn, Fe, nitrogen and phytate levels in 24 teff varieties. Nitrogen levels (A). Correlation between Zn and nitrogen content (B). Correlation between Fe and nitrogen content (C). Phytate levels (D). Correlation between Zn and phytate content (E). Correlation between Fe and phytate content (F).

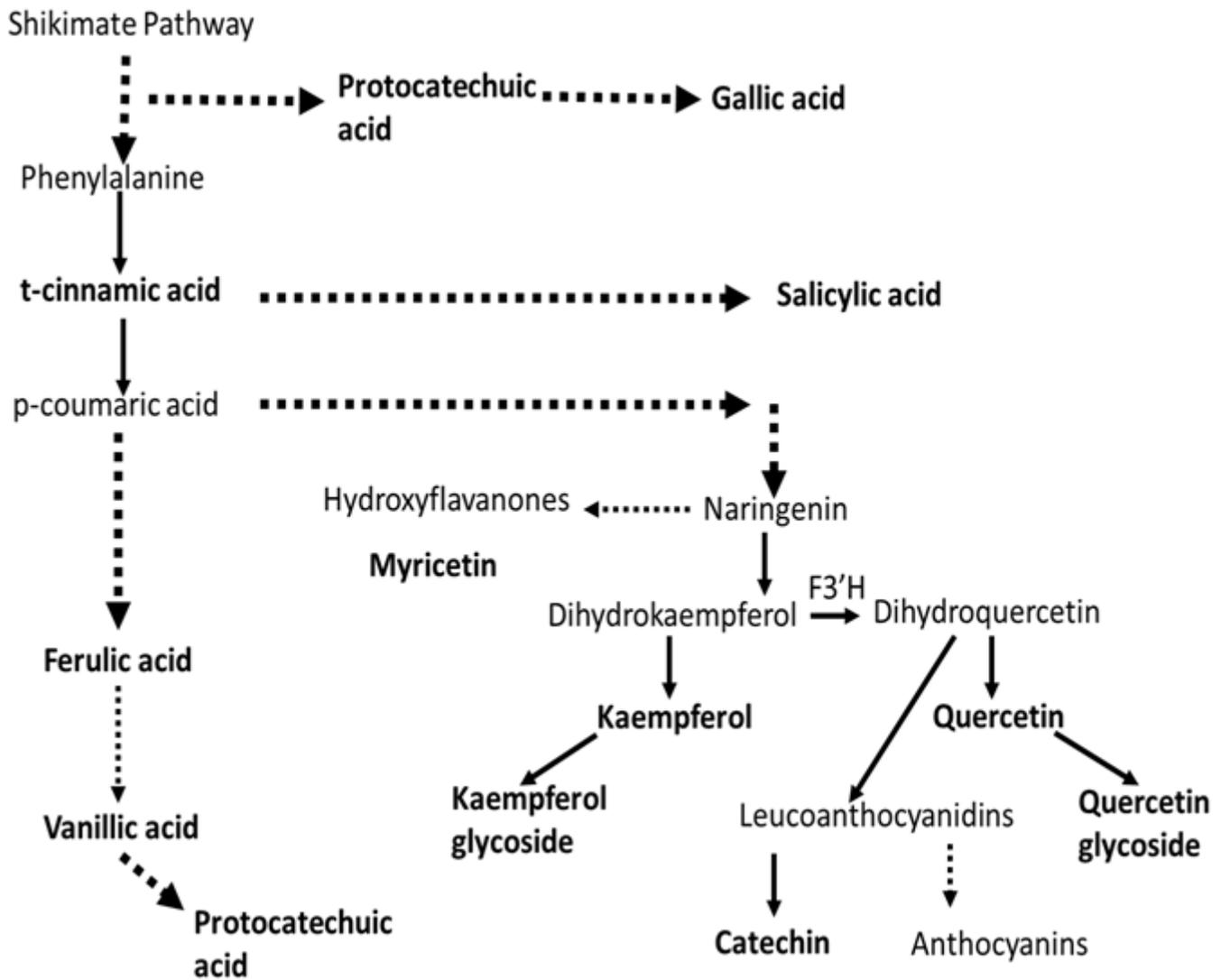


Figure 3

Biochemical pathway of selected phenolic compounds. The schematic shows the biochemical pathway relationship between the phenolic compounds measured in this study. Flavonoid 3'-hydroxylase (F3'H) is a key enzyme in the conversion of dihydrokaempferol to dihydroquercetin.

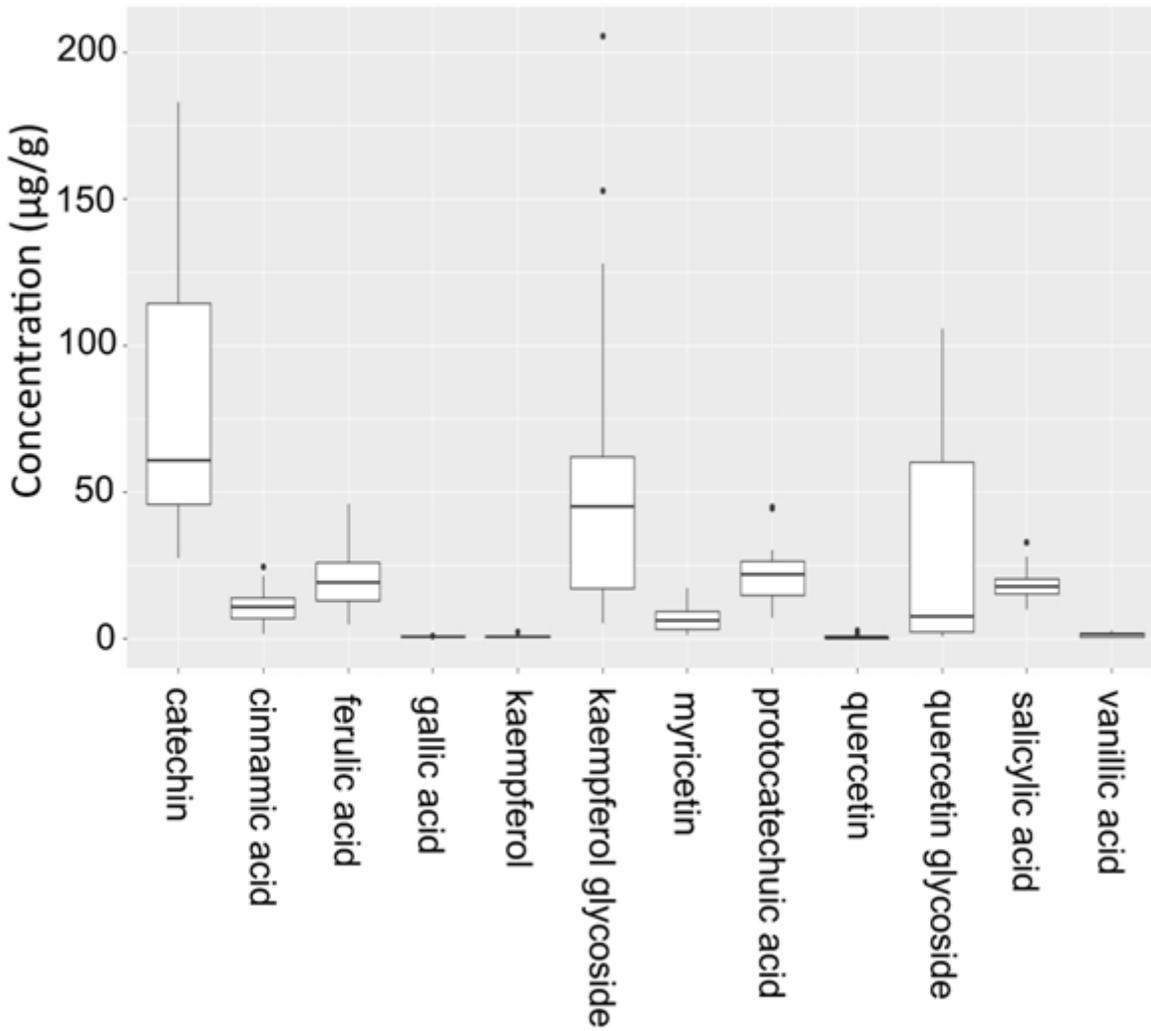


Figure 4

Levels of twelve phenols identified in the flours of 24 teff varieties. Each box shows the range of concentrations in µg/g of 12 phenolic compounds measured in the 24 teff varieties. The box defines the upper and lower quartiles. The lines extending vertically, “whiskers”, indicate variability outside the upper and lower quartiles, with points outside those lines, represented by dots, being considered outliers.

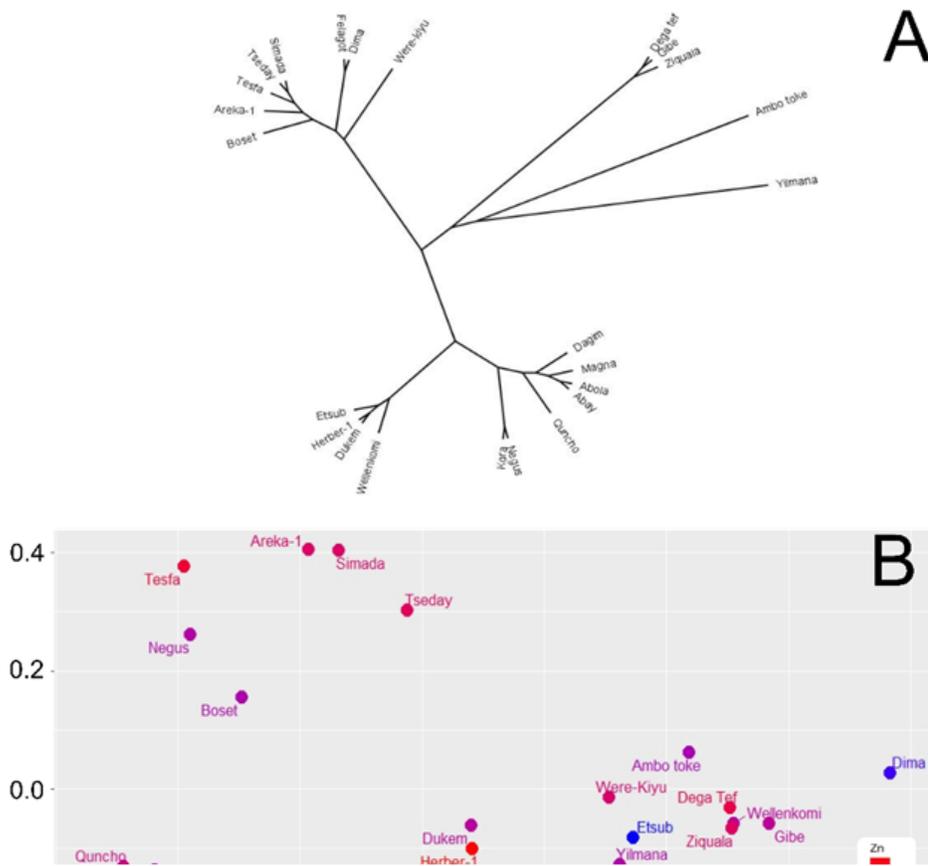


Figure 5

Population characteristics of the 24 sequenced teff varieties constructed using genome wide SNPs. (A) Phylogenetic tree, (B) PCA plot overlaid with grain Zn content, and (C) Population structure predicted using FastStructure.

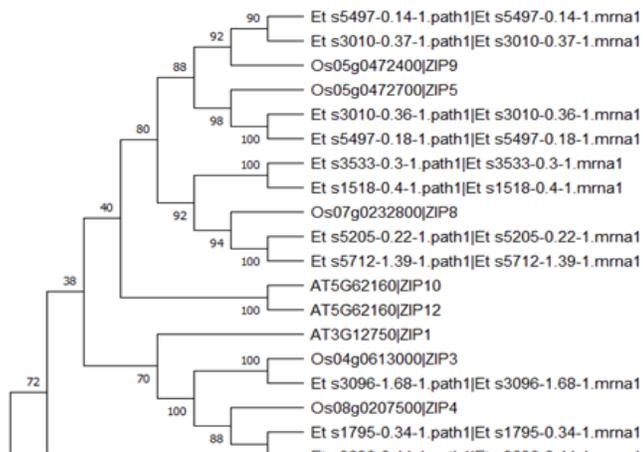


Figure 6

Phylogenetic tree of ZIP transporters. Shown is a maximum likelihood tree of Arabidopsis (At), rice (Os) and teff(Et) ZIP amino acid sequences. Values at each node are the calculated boot strap values from 50 iterations of the tree.

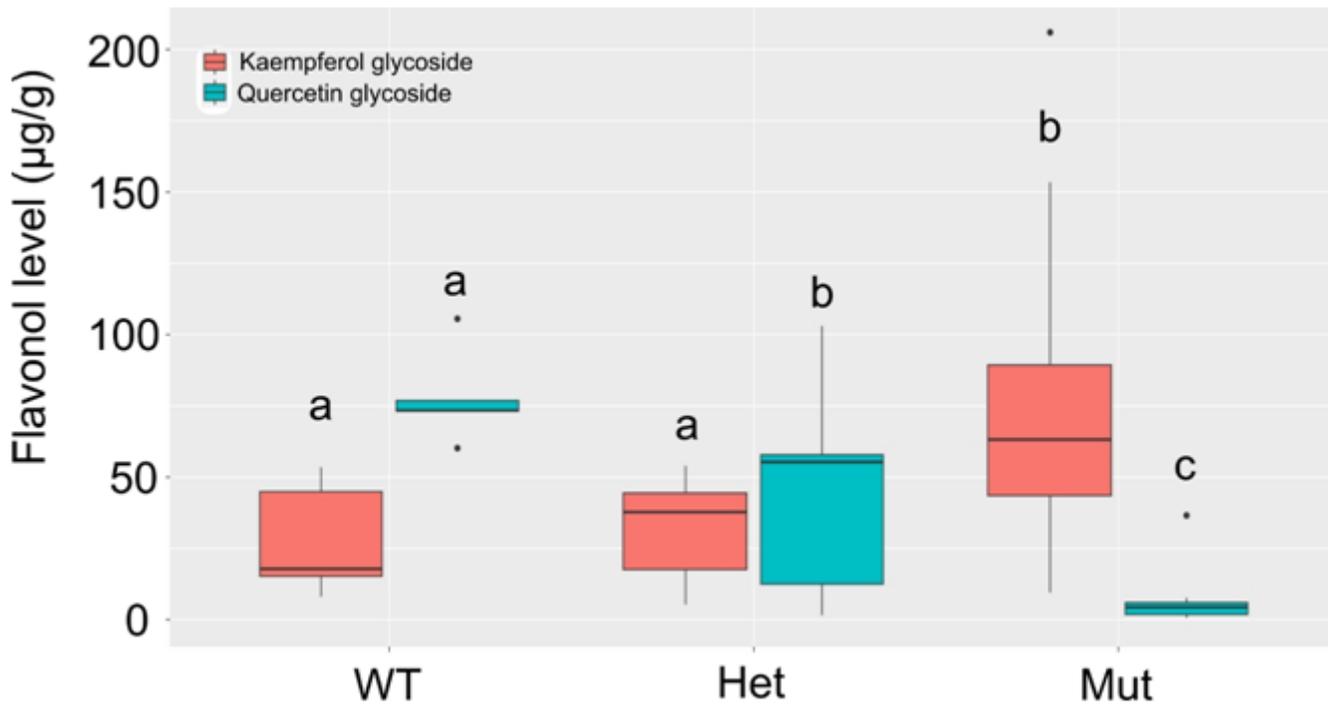


Figure 8

Haplotype analysis of a F3'H gene in teff. Kaempferol glycoside and quercetin glycoside levels in teff flours given the genotype of the T G mutation in the second intron of Et_s3159-0.29-1.mrna1. A different letter represents a significant difference (p val<0.05) between genotypes for either kaempferol glycoside or quercetin glycoside.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplFiguresfinal.docx](#)
- [Supltables.xlsx](#)