



OPEN ACCESS

EDITED BY

Leonardo Alfredo Ornella,
Cubiqfoods SL, Spain

REVIEWED BY

Karansher Singh Sandhu,
Bayer Crop Science, United States
Nikwan Shariatipour,
Shiraz University, Iran
Leonardo Caproni,
Sant'Anna School of Advanced Studies, Italy

*CORRESPONDENCE

Karen Cichy,
✉ karen.cichy@usda.gov

RECEIVED 30 October 2023

ACCEPTED 03 January 2024

PUBLISHED 06 February 2024

CITATION

Izquierdo P, Sadohara R, Wiesinger J, Glahn R,
Urrea C and Cichy K (2024), Genome-wide
association and genomic prediction for iron and
zinc concentration and iron bioavailability in a
collection of yellow dry beans.
Front. Genet. 15:1330361.
doi: 10.3389/fgene.2024.1330361

COPYRIGHT

© 2024 Izquierdo, Sadohara, Wiesinger, Glahn,
Urrea and Cichy. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Genome-wide association and genomic prediction for iron and zinc concentration and iron bioavailability in a collection of yellow dry beans

Paulo Izquierdo¹, Rie Sadohara¹, Jason Wiesinger²,
Raymond Glahn², Carlos Urrea³ and Karen Cichy^{1,4*}

¹Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States, ²USDA-ARS, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, United States, ³Department of Agronomy and Horticulture, Panhandle Research and Extension Center, University of Nebraska-Lincoln, Scottsbluff, NE, United States, ⁴USDA-ARS, Sugarbeet and Bean Research Unit, East Lansing, MI, United States

Dry bean is a nutrient-dense food targeted in biofortification programs to increase seed iron and zinc levels. The underlying assumption of breeding for higher mineral content is that enhanced iron and zinc levels will deliver health benefits to the consumers of these biofortified foods. This study characterized a diversity panel of 275 genotypes comprising the Yellow Bean Collection (YBC) for seed Fe and Zn concentration, Fe bioavailability (FeBio), and seed yield across 2 years in two field locations. The genetic architecture of each trait was elucidated via genome-wide association studies (GWAS) and the efficacy of genomic prediction (GP) was assessed. Moreover, 82 yellow breeding lines were evaluated for seed Fe and Zn concentrations as well as seed yield, serving as a prediction set for GP models. Large phenotypic variability was identified in all traits evaluated, and variations of up to 2.8 and 13.7-fold were observed for Fe concentration and FeBio, respectively. Prediction accuracies in the YBC ranged from a low of 0.12 for Fe concentration, to a high of 0.72 for FeBio, and an accuracy improvement of 0.03 was observed when a QTN, identified through GWAS, was used as a fixed effect for FeBio. This study provides evidence of the lack of correlation between FeBio estimated *in vitro* and Fe concentration and highlights the potential of GP in accurately predicting FeBio in yellow beans, offering a cost-effective alternative to the traditional assessment of using Caco2 cell methodologies.

KEYWORDS

biofortification, iron bioavailability, GWAS-assisted genomic prediction, *Phaseolus vulgaris* L, diversity panel

Introduction

Dry bean is the most important legume for human consumption worldwide (FAO et al., 2022) providing high levels of protein, dietary fiber, and micronutrients such as iron (Fe) and zinc (Zn) (Uebersax et al., 2022). Dry beans are also an appealing crop choice for biofortification because of their low environmental footprint, ability for symbiotic nitrogen fixation, and long shelf life that minimizes food waste (Willett et al., 2019). Biofortification

initiatives since 2005 have focused on increasing dry bean Fe and Zn concentrations through breeding as a means to improve the nutritional status of humans with Fe and Zn deficiencies (Beebe, 2020). Iron concentrations in beans are projected to decrease in climate change scenarios of rising CO₂ (Smith et al., 2017). Combatting climate change necessitates a shift toward plant-based diets, which, though beneficial for the environment, typically offer lower iron bioavailability compared to animal-based alternatives (van Wonderen et al., 2023). This transition is likely to exacerbate human iron deficiency issues in the future (MacDiarmid and Whybrow, 2019). Consequently, emphasizing iron and its bioavailability in bean breeding programs becomes crucial to mitigate this nutritional concern.

As bean breeders have worked for the last 2 decades to biofortify beans by increasing minerals concentrations, major efforts have been made to understand the genetic architecture of seed Fe and Zn accumulation (Cichy et al., 2009; Blair et al., 2011; Blair and Izquierdo, 2012; Blair et al., 2013; Mahajan et al., 2017; Caproni et al., 2020; Delfini et al., 2021; Gunjača et al., 2021; Cichy et al., 2022; Nazir et al., 2022). A meta-QTL (MQTL) analysis identified 12 stable MQTLs over different genetic backgrounds and environments (Izquierdo et al., 2018). While there has been a push towards breeding for increased seed micronutrient concentration, multiple studies have also underscored the value of breeding for enhanced Fe bioavailability (Katuuramu et al., 2018; Glahn et al., 2020; Katuuramu et al., 2021). These studies have indicated that Fe concentration and Fe bioavailability are not necessarily positively correlated. Increased Fe levels are often associated with increased amounts of polyphenols and phytate in the seed and both compounds form complexes with Fe that can pass through the human small intestine undigested, thereby negating the potential benefit of the higher Fe concentration (Tako et al., 2015).

Seed Fe and Zn concentrations are quantitative traits controlled by many loci across the genome in dry beans (Izquierdo et al., 2018) and are strongly influenced by the environment (Katuuramu et al., 2021). Although Fe bioavailability is also a quantitative trait, a study reported across nine locations in Uganda found it to be stable across environments (Katuuramu et al., 2021). The Caco-2 cell bioassay is the most cost-effective and practical method for assessment of Fe bioavailability from foods; however, the model requires highly trained technicians and laboratory resources that are often out of reach for most breeding programs (Glahn et al., 1998). One study in common bean has dissected the genetic architecture of Fe bioavailability through genome-wide association studies (GWAS), identifying five SNP associations distributed on chromosomes 6, 7, and 11 with phenotypic variability explained by the associated markers ranging from 8% to 13% (Katuuramu et al., 2018).

Genomic prediction uses genotype and phenotype from training datasets to estimate the phenotypes of new lines within a testing dataset, thereby reducing the need for extensive phenotyping and increasing selection intensity (Bernardo, 2020). This approach predicts phenotypes using all genetic markers collectively (Meuwissen et al., 2001), and the accuracy, often represented as the Pearson correlation of observed vs predicted phenotypes, depends on several factors. These include the heritability of the trait, linkage disequilibrium (LD) with causal loci, and the size of the population. The Genomic Best Linear Unbiased Predictor (GBLUP) is a widely employed parametric linear model that captures the

additive relationships between lines. However, semiparametric models like the Reproducing Kernel Hilbert Space (RKHS) can potentially identify non-additive effects, thereby potentially enhancing prediction accuracy (Gota and Gianola, 2014; Cuevas et al., 2016). Another critical aspect of genomic prediction is the genetic relationship between the training and testing datasets (Voss-Fels et al., 2019). The composition of the training data set and its relationship with the new lines to be predicted are crucial for maximizing prediction accuracy (Isidro et al., 2021; de los Campos et al., 2013). While several approaches have been proposed to optimize the training population, most of them have the assumption that one training is optimal for all individuals in the testing dataset (Lopez-Cruz et al., 2022). Considering that a high level of genetic heterogeneity is plausible in breeding programs, using only one optimal population may include individuals distantly related to the individuals in the testing dataset, reducing predictive ability (Lorenz and Smith, 2015). Recently, a sparse selection index (SSI) was proposed to identify a training set for each individual in the testing set (Lopez-Cruz and de los Campos, 2021). The use of this approach has increased prediction ability in multigeneration data up to 10% and 17% in wheat and maize, respectively, compared to the GBLUP (Lopez-Cruz et al., 2021; Lopez-Cruz et al., 2022).

Large genetic variability exists for Fe and Zn concentration, as well as for Fe bioavailability, both among and within market classes (Katuuramu et al., 2021). This variability provides evidence that achieving genetic gain in these traits is feasible. However, due to the complex genetic architecture of these traits, the use of genomics approaches is necessary to identify causative alleles and to evaluate prediction models. These approaches are essential for enhancing genetic gain in these traits, thereby increasing the nutritional value of new cultivars. Furthermore, there is a lack of information in the literature about the performance of prediction models trained on diversity panels when implemented in breeding lines. To address these challenges, we utilized Single Nucleotide Polymorphism (SNP) markers to identify genomic regions associated with Fe and Zn concentration and Fe bioavailability in a Yellow Bean Collection (YBC). Additionally, we evaluated the prediction accuracy of genomic prediction in the YBC and investigated whether the inclusion of SNP markers as fixed effects improves prediction accuracy. Finally, we assessed the predictive ability of models trained using the YBC on advanced yellow lines, exploring the potential of using genomic prediction to enhance nutritional quality in the yellow bean market class.

Materials and methods

Plant material

The yellow bean collection (YBC) comprised of 275 *Phaseolus vulgaris* L. accessions was grown at the Michigan State University Montcalm Research Farm in Entrican, MI, and at University of Nebraska field sites in Scottsbluff and Mitchell, Nebraska, during 2018 and 2019 growing seasons, respectively. A detailed description of each accession in the YBC is provided by Sadohara et al. (2022). In all years and locations, the YBC was grown in a randomized complete block design with two

replications. Additionally, 82 F5 yellow bean breeding lines were planted at the Montcalm Research Farm in 2019. These lines originated from biparental crosses involving seven Andean accessions, six of which are part of the YBC, as parents. These accessions were chosen for breeding due to their agronomic characteristics, stable yield across seasons, desirable seed color, short cooking times, and high iron bioavailability. The YBC planted in both years was used as the training set, while the 82 lines evaluated in 2019 were used as the prediction set to simulate the implementation of GP using a pre-breeding population. The local standard agricultural practices were followed for research plot scale dry bean production (Sadohara et al., 2022). Seed weight (SW) and yield (YD) were collected after harvest. SW (g) was obtained from weighing 100 seeds, and YD (kg ha⁻¹) was calculated based on the plot size and corrected to the moisture content in the seed of 18%.

Mineral concentrations (Michigan and Nebraska)

The protocol described by Glahn et al. (2020) was used to measure Fe and Zn concentration from raw seeds in the YBC grown in Michigan (MI) and Nebraska (NE) in 2018 and 2019 and the breeding lines grown in MI in 2019. Briefly, seeds were rinsed and cleaned with distilled water to remove dust and debris. The cleaned seeds were lyophilized and milled, and 0.50 g of the powder was predigested in boro-silicate glass tubes with concentrated ultrapure nitric acid and perchloric acid mixture (60:40 v/v) for 16 h at room temperature. Samples were then placed in a digestion block and heated incrementally over 4 h to a temperature of 120°C with refluxing. After incubating at 120°C, ultrapure nitric acid was subsequently added to each sample before raising the digestion block temperature to 145°C for an additional 2 hr. The temperature was then raised to 190°C for 10 min to evaporate remaining acid before cooling to room temperature. Digested sample was re-suspended in 20 mL of ultra-pure water before analysis using ICP-AES (inductively couple plasma atomic emission spectrometry; Thermo iCAP 6500 Series, Thermo Scientific, Cambridge, United Kingdom) with quality control standards following every 10 samples. All samples were digested and measured with 0.5 µg/mL of Yttrium (final concentration) purchased from High Purity Standards (10 M67-1) as an internal standard to ensure batch-to batch accuracy.

Fe bioavailability (Michigan only)

The YBC grown in MI in 2018 and 2019 was cooked using an automated Mattson cooker, as reported previously (Sadohara et al., 2022). The cooked samples were lyophilized and milled, and 0.50-g of powder was subjected to an *in vitro* digestion/Caco-2 cell culture model for the determination of Fe bioavailability, as described previously by Glahn et al. (1998). Fe uptake is measured as the increase in Caco-2 cell ferritin production (ng ferritin per milligram of total cell protein) after exposure to simulated gastric and intestinal digest. Fe bioavailability is expressed as a percentage score of Caco-2 cell ferritin formation that is relative to

a cooked/lyophilized/milled white kidney bean (Snowdon). The white kidney bean reference control was run with each bioassay to index the ferritin/total cell protein ratios of the Caco-2 cells over the course of multi-year experiment. The Snowdon white kidney bean was used as a reference control because this cultivar is commercially produced in North America and has high Fe bioavailability due to the lack of polyphenols that inhibit the absorption of iron (Wiesinger et al., 2019). The mean ferritin formation values of Snowdon across assays were 19.71 and 15.78 ng ferritin/mg total cell protein in 2018 and 2019, respectively. Fe bioavailability was not measured in the 82 breeding lines.

Statistical analyses

The rows and columns from the field were used as random effects to fit a linear mixed model for SW and YD using the functions “SpATS” and “SAP” for the R package SpATS (Rodríguez-Álvarez et al., 2018). The effect of the genotype on the phenotype was fitted as fixed to obtain the best linear unbiased estimator (BLUE) of SW and YD. The mineral concentrations were collected in one field replication in Michigan and Nebraska during two growing seasons (2018–2019), while Fe bioavailability was collected in one field replication only in Michigan during two growing seasons (2018–2019). The mean of two technical replications was used as the mineral concentration and Fe bioavailability for each environment and year. The variance component analysis was conducted using the R package statGenGxE, adopting a factorial structure of locations per year (StatGenGxE, 2023). In this mixed linear model, terms for year, location and location:year were treated as fixed effects, while the effects of genotype, genotype:year, and genotype:location were treated as random.

Genotypic data

DNA was extracted from trifoliate leaves using NucleoSpin Plant II Kit (Macherey–Nagel, Duren, Germany) following the “Genomic DNA from plants” protocol as described previously by Sadohara et al. (2022). DNA concentration was measured using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, Waltham, MA, United States), and 10 ng/µL of DNA was used for GBS library preparation with a single restriction enzyme, ApeKI. Each plate of 96-wells was sequenced in a lane of an Illumina HiSeq platform using single-end reads. The libraries were demultiplexed using NGSEP (v3.1.2) (Tello et al., 2019). Adapters and low-quality bases from the raw sequencing data were trimmed using Cutadapt v 1.16, and the processed reads were aligned to the reference genome of *P. vulgaris* v2.1 G19833 (Schmutz et al., 2014) using Bowtie2 (v2.2.30) (Langmead and Salzberg, 2012) with default parameters. The SNP calling was carried out by using NGSEP software following the recommended parameters for GBS data (Perea et al., 2016). The merged genotypic matrix was filtered with NGSEP for variants that were in the predicted repetitive regions of the common bean genome (Lobaton et al., 2018), non-biallelic, genotype quality above 30, a maximum observed heterozygosity of 0.05 per SNP, more than 50% of missing data per site, and minor allele frequency (MAF) above

0.05. Besides, SNPs in linkage disequilibrium above 0.9 using a window of 500 SNPs were removed using Bcftools (Li, 2011). The resulting genotype matrix was imputed using Beagle V5.4 with default parameters (Browning et al., 2018).

Genome-wide association study (GWAS)

The GWAS was conducted using the Fixed and Random Model Circulating Probability Unification (FarmCPU) method, as implemented in GAPIT v3 (Wang and Zhang, 2021). The YBC accessions from each year and location were used for this analysis. FarmCPU iteratively adjusts for fixed (individual SNP markers) and random effects (polygenic background and population structure), enhancing true association detection. A Shapiro-Wilk test was carried out on the phenotypic data, and traits not considered normal (p -value < 0.05) were transformed using the rank-based inverse normal transformation (INT). Subsequently, GWAS was conducted on both transformed and untransformed phenotypes. Significant associations were identified using a Bonferroni-corrected threshold of $\alpha = 0.05$.

Genomic prediction

Four prediction models were assessed in the YBC: The Reproducing Kernel Hilbert Space (RKHS) regression, sparse selection indices (SSI), and RKHS and SSI using QTN identified through GWAS as fixed variables. The RKHS regression was implemented using the kernel averaging (KA) approach (de los Campos et al., 2010) with kernels estimated using extreme bandwidth parameters (0.2, 1, 5) (González-Camacho et al., 2012). The Gaussian kernels utilize a positive-definite kernel represented by: $K = \exp(-\theta d_{ij}^2)$, where K represents the kernel, θ is the bandwidth parameter, and d_{ij}^2 represents the scaled squared Euclidean distance between individuals i and j based on their SNPs (Lopez-Cruz et al., 2021).

The SSI was obtained by imposing an L1-penalty on a selection index using additive genomic relationships (GSSI) (Lopez-Cruz and de los Campos, 2021). To optimize the penalization value λ in SSI, 10-fold cross validation was carried out in the training subset, and SSI was derived over a grid of 100 values of λ . The accuracy measured as the Pearson correlation between SSI and each trait was used to identify the value of λ that maximized accuracy in each cross validation. The optimal value of λ was defined as the average value of λ that maximized accuracy across each cross-validation and was used to predict the validation subset. The training-validation procedure described above for KA and SSI was repeated 100 times by randomly assigning samples from the training set into training and validation subsets and was implemented using the R packages SFSI and BGLR (Lopez-Cruz et al., 2020; Pérez-Rodríguez and de Los Campos, 2022). The prediction ability for all models is expressed as a Pearson correlation coefficient between the observed and predicted values in the validation subset.

Prediction set

To simulate the implementation of GP, 82 breeding lines from 2019 were included as the prediction set, as previously described.

The YBC was employed to train the KA and SSI models. Given that the breeding lines are derived from crosses between Andean accessions, the models were trained with two distinct YBC datasets: one encompassing all accessions and another restricted to Andean accessions. Additionally, to train the models, different proportions of individuals of each bi-parental family (0%, 10%, 20%, 30%) were randomly assigned to the training sets. The procedure described above was repeated 100 times and was implemented using the R packages SFSI and BGLR (Lopez-Cruz et al., 2020; Pérez-Rodríguez and de Los Campos, 2022).

Results

Phenotypic data

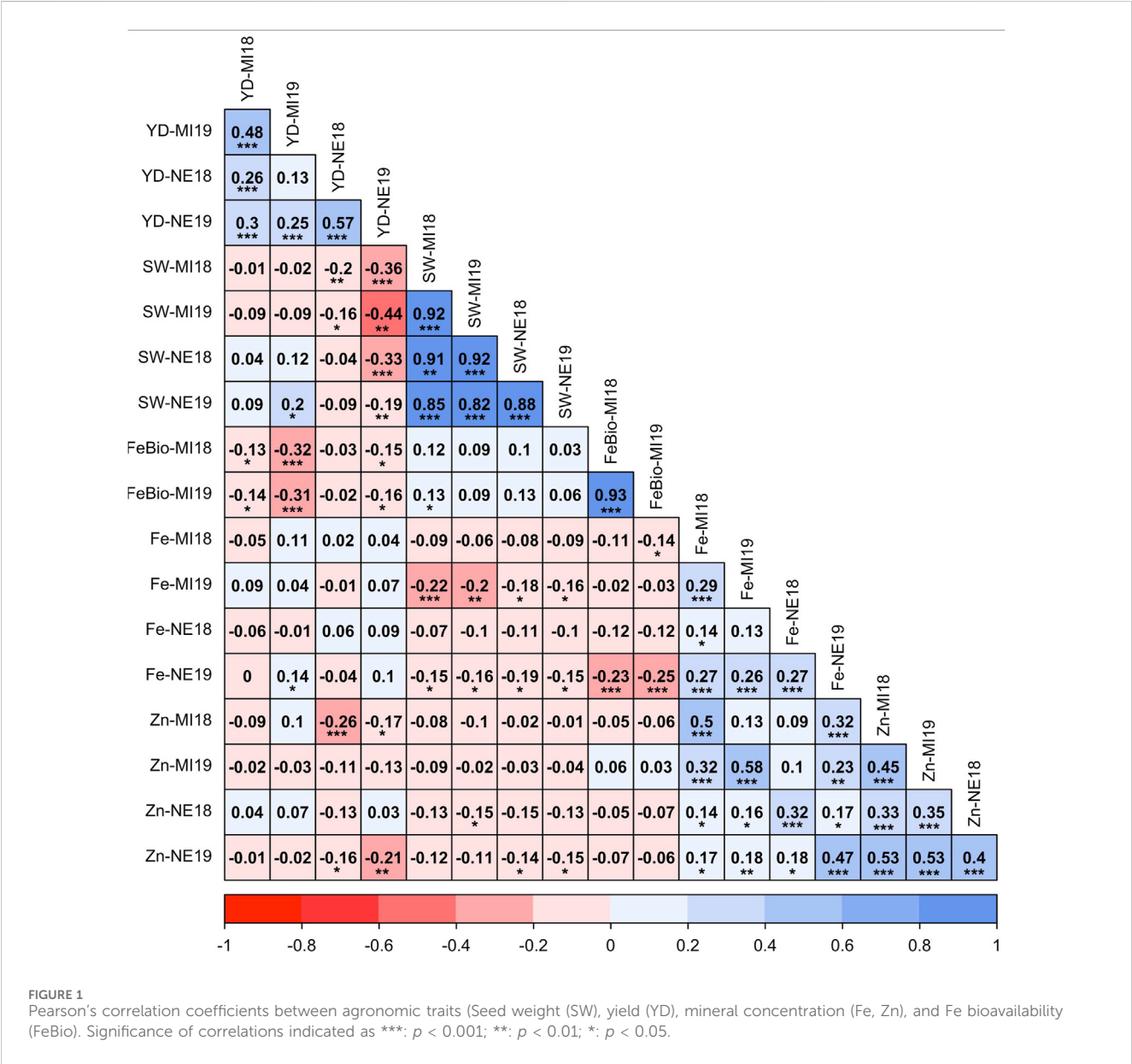
The YBC was grown in field locations in MI and NE in 2018 and 2019 and was evaluated for seed yield, seed weight, Fe and Zn concentration in MI and NE 2018–2019, and FeBio was evaluated in Michigan 2018–2019. The YBC exhibits large variability for agronomic and nutritional quality traits (Table 1 and Supplementary Figures S1, S2). The FeBio distribution was bimodal in the two growing seasons measured in Michigan. The concentration of Fe ranged from 43 to 118 $\mu\text{g/g}$ in Michigan 2018, 39–106 $\mu\text{g/g}$ in Michigan 2019, 28–76 $\mu\text{g/g}$ in Nebraska 2018, and 36–72 $\mu\text{g/g}$ in Nebraska 2019 (Table 1). Four genotypes had concentrations above 100 $\mu\text{g/g}$ in Michigan, three in 2018 (YBC050, YBC136, and YBC279) and one in 2019 (YBC171). The concentration of Zn ranged from 20 to 41 $\mu\text{g/g}$ in Michigan 2018, 17–41 $\mu\text{g/g}$ in Michigan 2019, 19–37 $\mu\text{g/g}$ in Nebraska 2018, and 21–43 $\mu\text{g/g}$ in Nebraska 2019 (Table 1). FeBio values in Michigan varied from 11% to 151% in 2018 and from 13% to 118% in 2019, compared to the reference control Snowdon (Table 1). Twenty-seven genotypes had higher FeBio values than the high FeBio check variety, Snowdon, in 2018 and 2019.

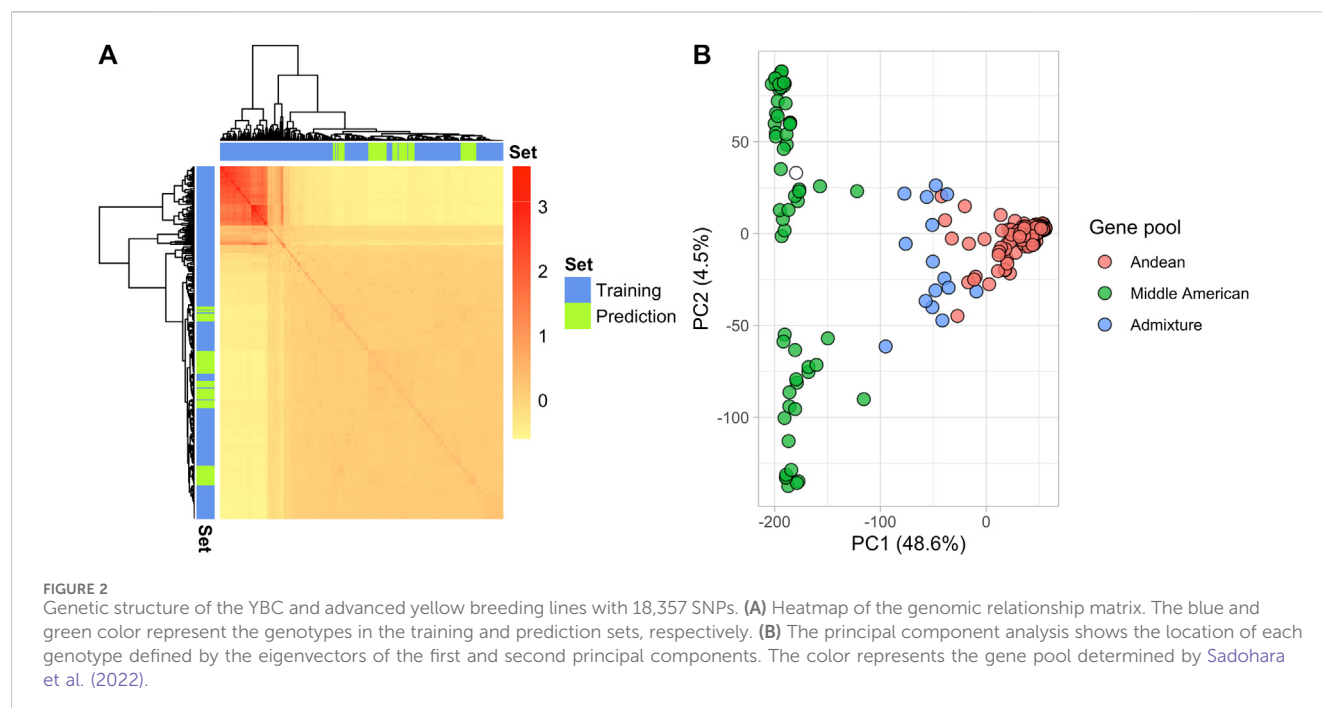
The variance component analysis revealed a significant effect of genotype on the agronomic and nutritional quality traits assessed in the current study (Supplementary Table S1). The most important source of variation for Fe concentration was the location followed by the Year:Location, while genotype explained most of the variation in Zn concentration and FeBio. Indeed, the genotype explained 93%, 5%, and 37% of the phenotypic variation in FeBio, Fe, and Zn concentration, respectively. The large effect of the genotype on the variability of FeBio is supported by the high h^2 of FeBio found in 2018 (0.85) and 2019 (0.91).

Despite the differences between location and years, the measurements of all traits were positively correlated across years (Figure 1). The correlation between Fe concentration and yield was not significant in any environment, ranging from -0.05 in Michigan in 2018 to 0.10 in Nebraska in 2019, while the correlation between yield and Zn concentration was negative in both locations and years. Fe and Zn concentrations were positively correlated, ranging from $r = 0.32$ ($p < 0.001$) in Nebraska in 2018 to $r = 0.58$ ($p < 0.001$) in Michigan in 2019. Both Fe and Zn concentrations were negatively correlated with seed weight in both locations and years. The most stable trait across years (2018–2019) was FeBio, with a correlation between years of 0.93 ($p < 0.001$) in Michigan. FeBio was negatively correlated with yield in 2018 ($r = -0.13$, $p = 0.05$) and 2019 ($r = -0.31$, $p = 0.001$).

TABLE 1 The mean and the range of yield, seed weight, Fe bioavailability, Fe and Zn concentration of the YBC grown in MI and NE in 2018 and 2019.

Trait	MI 2018		MI 2019		NE 2018		NE 2019	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Yield (kg/ha)	2111 (394)	1171–3299	1649 (477)	536–3065	1974 (647)	653–5,275	686 (372)	52–2124
SW (g)	38 (8)	18–60	40 (10)	18–73	35 (7)	15–54	30 (1)	26–35
Fe (µg/g)	69 (12)	43–118	65 (9)	39–106	45 (9)	28–76	51 (6)	36–72
Zn (µg/g)	29 (4)	20–41	28 (4)	17–41	25 (3)	18–37	27 (3)	21–43
FeBio (% of control)	54 (31)	11–151	51 (30)	13–118	-	-	-	-





Population structure

In total, 18,357 SNP markers were retained after filtering in the 275 YBC and 82 breeding lines. The genomic relation matrix revealed two groups with strong genetic relationships, representing the germplasm from the two gene pools in *P. vulgaris* (Figure 2A). As expected, the breeding lines included in 2019 were clustered in the same gene pool as their founders (Andean) (Figure 2A). Genetic diversity was evaluated through principal component analysis (PCA). The first two principal components explained 53.1% of the variance (Figure 2B). PC1 separated the Andean and Middle American gene pools as reported by Sadohara et al. (2022). However, the variance explained by PC1 (48.6%) in this study differs from the PC1 (63.8%) reported by Sadohara et al. This difference is due to different filters applied to the genotype matrix and the addition of the 82 breeding lines in the present study. Additionally, a Principal Component Analysis (PCA) conducted on the Andean accessions revealed that the parents used in the biparental crosses are closely related genetically (Supplementary Figure S3).

GWAS

Marker-trait associations were evaluated using the multi-marker approach, FarmCPU. In total, 24 SNP associations were identified in the untransformed phenotypic data, and 14 in the INT data, all with significance surpassing the threshold established by the Bonferroni correction (2.72×10^{-6}) (Supplementary Table S2). Three genomic regions on chromosomes 7, 10, and 11 were consistently associated with FeBio in Michigan across years using the untransformed phenotypes (Supplementary Figure S4). However, when analyzing the normalized, transformed phenotypes, these genomic regions on chromosomes 7, 10, and 11 showed associations in at least 1 year. In contrast, no consistent associations were identified for Fe and Zn

concentrations or yield between years or locations. Notably, the association with FeBio on chromosome 7 exhibited the highest SNP effect, represented as the difference between homozygous genotypes, over the two evaluated years. (Figure 3; Table 2).

Genomic prediction–training set

In total, 100 partitions were used to assess the prediction accuracy for the KA and SSI models in the training set composed of 275 YBC. The prediction accuracies of KA and SSI models for Fe and Zn concentrations and seed yield are presented in Figure 4 and Supplementary Table S3. The prediction accuracy was affected by trait heritability, and no significant differences were observed across models. KA models used all the genotypes in the training subset to train the model, while SSI used a penalization to select the individuals in the training subset to maximize the prediction ability. The average number across locations and years of individuals that support predictions using the SSI ranged from 34% to 95% of individuals in the training subset from Fe concentration and yield, respectively (Supplementary Table S3). The lowest prediction accuracy found in this study was detected in Fe concentration in Nebraska in 2018. Soil analysis revealed a pH of 8.0 at this location. The h^2 (0.05, 0.20) and prediction ability <0.17 for Fe and Zn concentration were likely affected by the alkaline soil of this location which causes reduced iron availability for plant uptake. Average prediction accuracy in 2018 and 2019 for yield ranged from 0.33 to 0.65 in Michigan and 0.51 to 0.65 in Nebraska, respectively (Figure 4).

Utilizing the 100 partitions previously described, the prediction accuracy of both the KA and SSI models was evaluated for FeBio prediction by incorporating the major QTN identified on chromosome 7 (Table 1). While there was no discernible difference in prediction accuracy between the KA and SSI models without QTN information, the inclusion of the major QTN

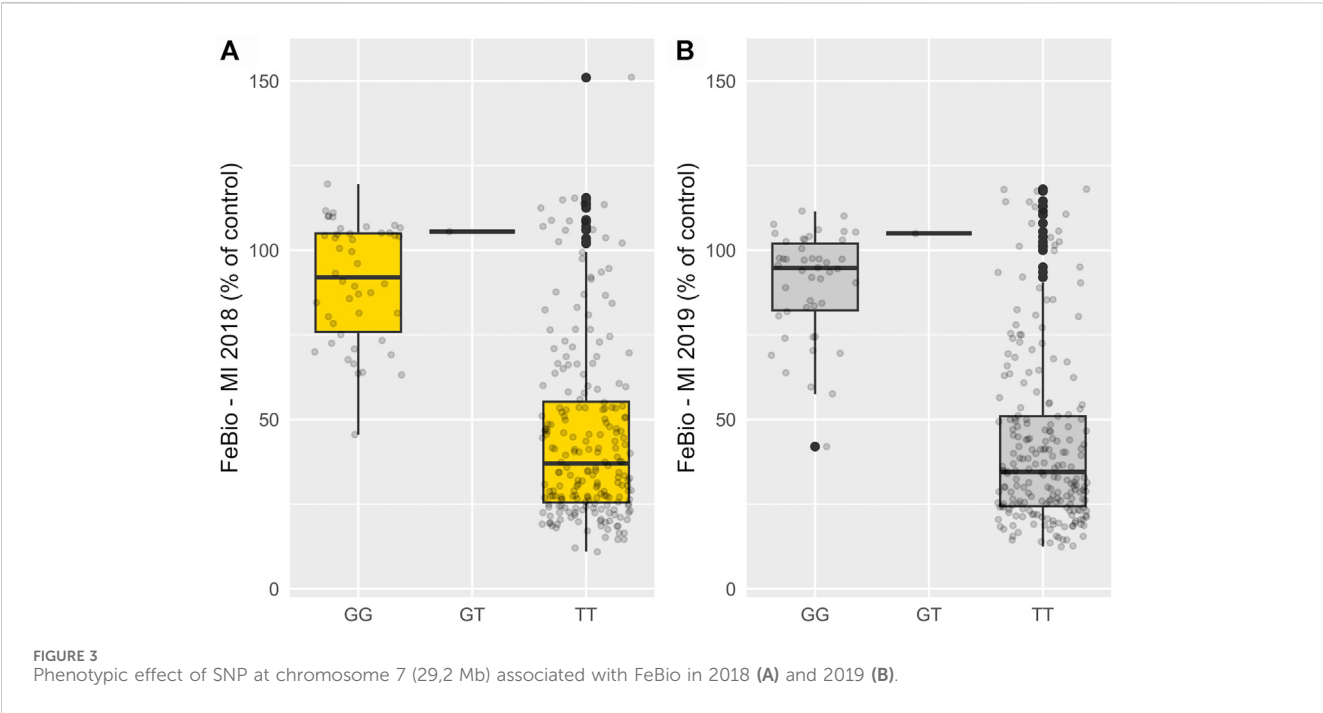


TABLE 2 QTN for FeBio identified by genome-wide association analysis over years using FarmCPU model.

Chr	Position	p-value	Year	SNP effect
7	29,169,848	4.43E-07	2018	45.2
7	29,169,848	1.75E-08	2019	46.9
10	42,422,135	5.90E-11	2018	3.3
10	42,422,135	2.67E-17	2019	3.9
11	3,626,904	4.75E-08	2018	9.3
11	3,626,904	6.77E-08	2019	8.0

enhanced the predictive capability of the KA model. Conversely, the SSI model's prediction accuracy remained unchanged with the addition of the QTN (Figure 5, Supplementary Table S3).

Genomic prediction–prediction set

To simulate the implementation of GP with a diversity panel where some lines were used for breeding, a set of 82 breeding lines was grown in Michigan in 2019 and used as a prediction set. The prediction accuracy of KA and SSI models, which used all the YBC accessions and included varying proportions (from 0% to 30%) of the prediction set for model training is presented in Figure 6 and Supplementary Table S4. For the traits measured in the prediction set (seed yield and Fe and Zn concentrations), the prediction accuracy of Fe concentration was not improved when breeding lines were added to train the model. However, the accuracy tended to increase for seed yield and Zn concentration by adding genotypes belonging to the same families in the prediction set. Additionally, the SSI model showed lower accuracy compared to KA for Fe and Zn concentration, and SSI resulted in higher predictions for seed yield

(Figure 6). Although the prediction accuracy obtained using SSI was overall lower compared with KA, the prediction is provided by, on average, <21% of the individuals used to train the model for Fe and Zn concentrations. For yield, the prediction is provided by, on average, <83% of the individuals used to train the model (Supplementary Table S4). Similar results were observed when only accessions from the Andean gene pool were used to train the models (Supplementary Table and Figure 5). However, prediction accuracy was consistently lower when the training population was reduced through the exclusion of non-Andean accessions.

Discussion

Biofortification is an important goal included in many breeding programs of dry beans. Three assumptions are used to release biofortified cultivars worldwide: *i*) Fe concentration is stable across environments, *ii*) the average Fe concentration in nonbiofortified dry beans is ~50 µg/g, *iii*) Fe bioavailability is positively correlated with Fe concentration (Glahn and Noh, 2021). In this study, we found that none of these assumptions are valid in the YBC. Although Fe concentration showed a positive correlation within and between locations in the two growing seasons, those correlations were $r < 0.3$. In both years, beans planted in Nebraska showed a lower Fe concentration (~48 µg/g) compared to Michigan (~67 µg/g), and the reasons for this difference may be attributed to the alkaline soil (pH = 8) found in the Nebraska location used in this study. Values of pH greater than 7 have been related to iron chlorosis and zinc deficiency in plants (Westfall and Bauder, 2011). Additionally, the Fe and Zn in soil available for plant uptake is related to the concentration of these microelements in seeds (Katuuramu et al., 2021). The complex genetic architecture of Fe and Zn concentration was described in

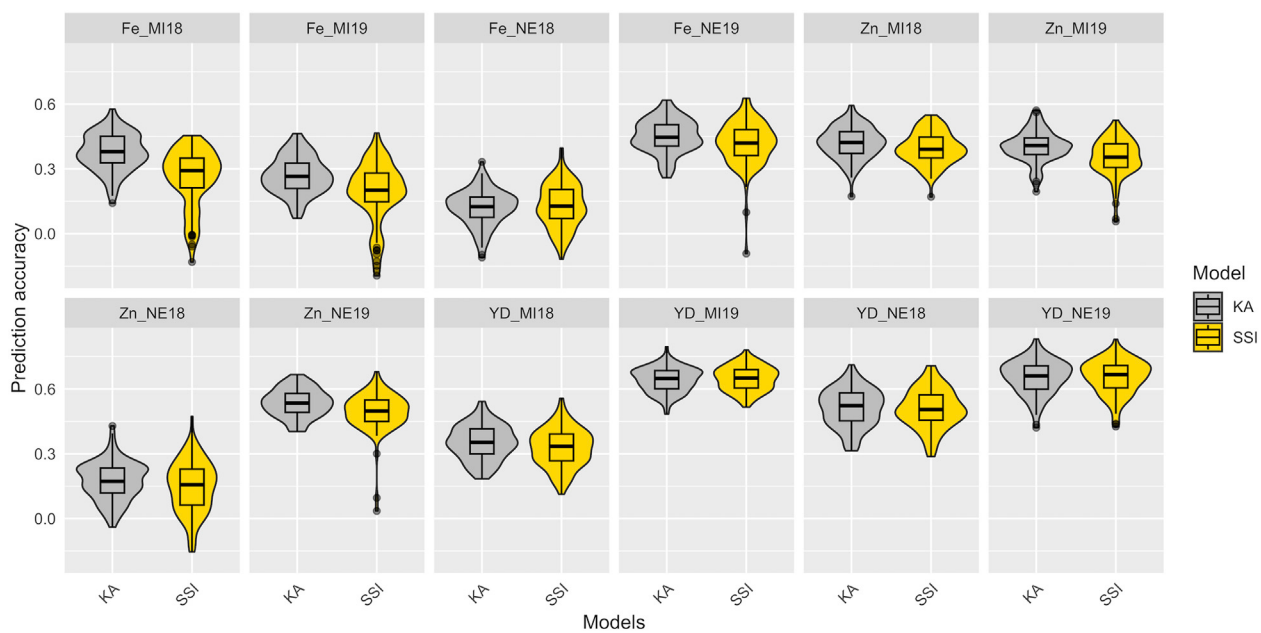


FIGURE 4
Prediction accuracy in the training data set of Fe-Zn concentration and yield using KA and SSI. The distribution of boxplots represents 100-fold cross-validations in the training set.

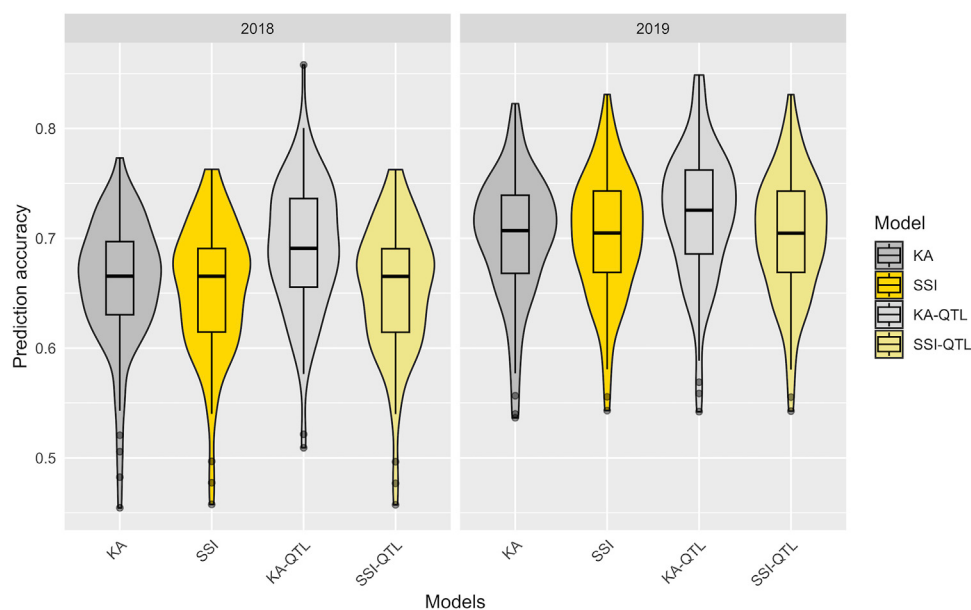
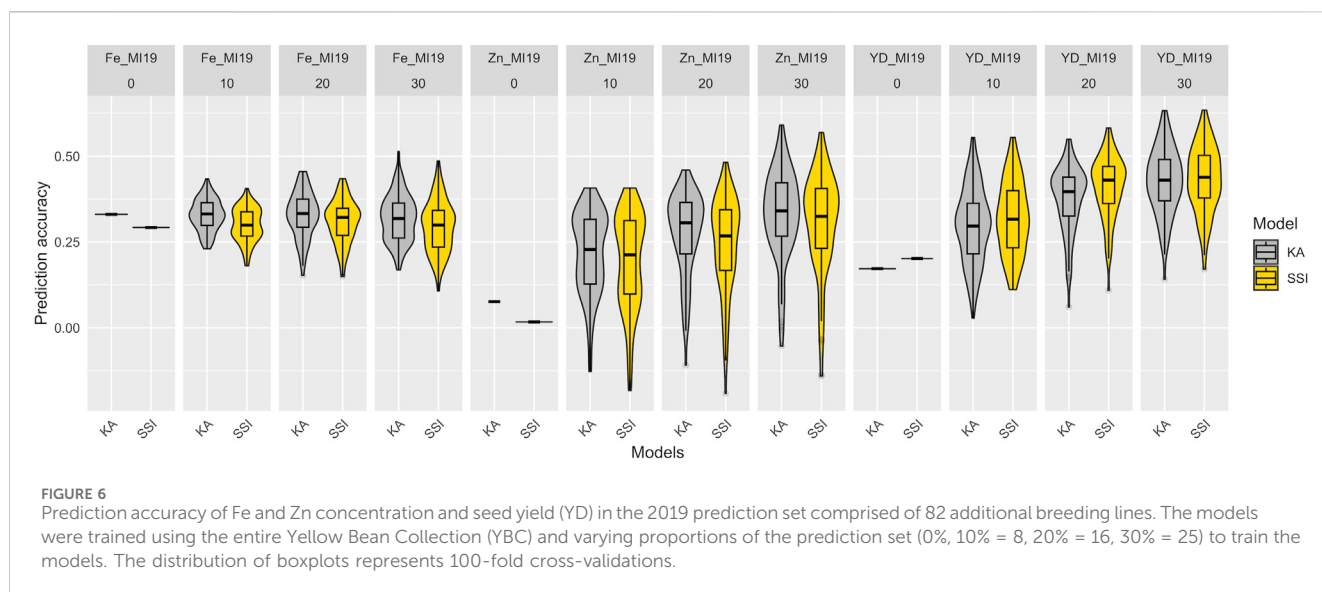


FIGURE 5
Prediction accuracy in the training data set of FeBio using KA, SSI, and QTN as fixed effect. The distribution of boxplots represents 100-fold cross-validations in the training set.

a meta-QTL analysis in seven populations of dry beans, and candidate genes related to the process of uptake, transport, and accumulation of these minerals were identified (Izquierdo et al., 2018). Due to the interactions with environmental factors and the complex genetic architecture of Fe and Zn concentrations, achieving a stable concentration of these microelements across environments

presents challenges. Our results found no correlation between Fe concentration and *in vitro* Fe bioavailability and similar results have been reported in dry beans (Katuuramu et al., 2018; Glahn et al., 2020; Katuuramu et al., 2021), which suggests that increases in the Fe concentration do not translate to higher FeBio in yellow beans. For this absence of relationship, Glahn & Noh (Glahn and Noh, 2021)



proposed to change the focus of biofortification from Fe concentration to FeBio in dry beans. There are two big advantages of breeding for FeBio over Fe concentration, higher heritability and delivery of absorbable Fe. We observed that FeBio was stable across years ($r = 0.93$), with a heritability >0.85 , which was 2-fold the heritability of Fe concentration presented for both growing seasons in Michigan.

The high phenotypic variability and heritability of FeBio in this study suggest that genetic improvement is feasible. However, we observed a strong relationship between seed type and FeBio, which could limit the genetic gain of FeBio in some market classes (Supplementary Figure S6). The seed types Manteca, Mayocoba, and white showed high FeBio values, and we consider that these three market classes should be used as priority in programs seeking to identify and develop varieties that deliver more bioavailable Fe. Interestingly, using data reported by Sadohara et al. (2021), we identified that lighter color seeds from the YBC tend to have shorter cooking times. Fast-cooking time has been associated with high FeBio and with more retention of nutrients in dry beans (Wiesinger et al., 2016), which means that the use of Manteca, Mayocoba, and white seed types have the potential to yield high FeBio and nutrients in fast-cooking genotypes, traits that are appealing for nutrition and customer acceptance.

GWAS

A strong environment effect was observed for yield, and Fe and Zn concentrations; and no consistent associations were identified for these traits on the YBC. The complex architecture of these traits and the GxE interaction has yielded hundreds of marker-trait associations across locations and populations (Izquierdo et al., 2018; Izquierdo et al., 2023), and the usage of strategies such as marker-assisted selection or GWAS-assisted genomic prediction does not appear promising for these traits (Keller et al., 2020).

FeBio exhibited large variability across different years and among individuals, with genetics playing a significant role in this variation. The bimodal distribution observed in FeBio likely stems

from the trait governed by major QTNs. A particular SNP on chromosome 7 at 29.2 Mb is associated with features such as light-colored seed, hilum ring, corona, and resistance to darkening in the YBC (Sadohara et al., 2021). This SNP is proximal to the P gene (*Phvul.007G171333*) at 28.8 Mb, a transcription factor essential for flavonoid biosynthesis. This gene's dominant allele, P, is necessary for seed coat color expression, while its recessive genotypes result in a white seed coat (McClean et al., 2018). This region on chromosome 7 was previously associated with FeBio in the Andean Diversity panel (Katuuramu et al., 2018). Furthermore, dark-colored bean seeds are observed to exhibit lower FeBio values (Wiesinger et al., 2018).

This finding further substantiates the relationship between seed color and FeBio. While the precise nature of the connection between seed color and FeBio remains elusive, several factors might explain it. Firstly, darker seeds typically have thicker seed coats. From ~6 to 39% of seed iron in dry beans is found in seed coats (Moraghan, 2004), a tissue known to contain high levels of antinutrients that can bind to and chelate iron (Ariza-Nieto et al., 2007). Given that *i*) the SNP effect on chromosome 7 (Table 2) was close to the population mean for FeBio in both years (Tables 1, 2), *ii*) the association was consistent across years, and *iii*) the association with FeBio has been reported in other genetic backgrounds, such as the Andean Diversity panel, this region might be indicative of a major QTN for FeBio.

Genomic prediction

In this study, we considered KA and SSI models to assess the accuracy of genomic prediction in the YBC. Across all traits, we found consistent results between the models, with prediction variations correlating with the heritability of each trait. While both models yielded comparable outcomes for the YBC, the specific individuals supporting the prediction of each line in the testing set differed. KA models used in each partition all the genotypes in the training subset, while SSI selected a particular set of support points as the optimal training set for each genotype in the testing set. The average number of supporting points selected by

the SSI model was related to the complexity of each trait, ranging from 34% (FeBio) to 95% (yield).

In a multi-generational wheat dataset spanning 8 years, the SSI model demonstrated superior prediction accuracy compared to GBLUP (Lopez-Cruz et al., 2022). As Lopez-Cruz and de los Campos (Lopez-Cruz and de los Campos, 2021) highlighted, SSI tends to achieve higher accuracies than GBLUP, especially in larger datasets. Nevertheless, the margin of increased prediction accuracy of SSI over GBLUP is typically slight (less than 0.05) and is strongly influenced by both the diversity of the dataset and the size of the training set. Although the YBC represents a diverse population that might benefit from SSI, its training set consists of fewer than 185 individuals. This is in contrast to the thousands of individuals used in the training models by Lopez-Cruz and de los Campos (2021) and Lopez-Cruz et al. (2022).

To assess the implementation of genomic prediction using the YBC and a subset of the Andean accession from the YBC as training populations, 82 breeding lines were used as the prediction dataset. By adding different percentages of each breeding line from the same family, prediction accuracy increased for seed yield and Zn. Moreover, SSI demonstrated slightly better performance than KA in predicting seed yield. The better performance of SSI compared to KA in seed yield could be the result of differences in allele frequencies and LD patterns in the families to be predicted compared to the YBC. Differences in allele frequencies and LD may lead to suboptimal estimation of breeding values using all sets of individuals to train the model (Lopez-Cruz et al., 2022). Overall, prediction accuracies were slightly lower when non-Andean accessions were excluded from the training dataset. This reduction in prediction accuracy could be attributed to the smaller size of the training population and suggests that the inclusion of Middle American accessions contribute to increased prediction accuracy in the prediction dataset.

The traits studied here exhibited polygenic inheritance, as evidenced by the continuous variation observed across locations and years. However, incorporating a major QTN associated with FeBio as a fixed variable in the KA model improved accuracy by an average of 0.03 (2018) and 0.02 (2019) compared to the model without fixed effects. It is worth noting that the improved predictive ability from the fixed marker in this study is likely related to the diversity in seed colors present in the YBC. Its positive impact might diminish in germplasm with less varied seed types. However, even without the inclusion of the fixed marker, the prediction accuracy for FeBio remained high (>0.65) using both KA and SSI models, suggesting the potential of genomic prediction for this trait in dry beans.

Conclusion

This study underscores the need to shift from the traditional biofortification strategy of merely increasing Fe concentration, advocating for developing dry bean cultivars with greater Fe bioavailability. The lack of correlation between Fe concentration and Fe bioavailability in this study aligns with findings from studies conducted in Africa and the United States (Katuuramu et al., 2018; Katuuramu et al., 2021), suggesting that reallocating

resources in breeding programs centered on biofortification would be beneficial. We discovered a robust association between Fe bioavailability and seed coat color, pinpointing that lighter seed colors like Manteca, Mayocoba, and White often exhibit higher Fe bioavailability. These should be prioritized to bolster the supply of bioavailable Fe for human consumption. While Fe bioavailability's significance is undeniable, its measurement might be out of reach for many breeding programs, and would thus require collaboration with laboratories capable of doing so. Our results indicated that genomic prediction offers high accuracy for Fe bioavailability, presenting a potential solution to mitigate the expenses and extensive duration associated with its measurement.

Data availability statement

The raw sequencing data for the dry bean individuals are available at NCBI under the accession number PRJNA1061170 (<https://www.ncbi.nlm.nih.gov/bioproject/1061170>). The GBS barcodes, along with the phenotype and genotype data for this study, can be found at <https://github.com/pauloizquierdo/GWASAssistedGenomicPrediction-YBC.git>.

Author contributions

PI: Investigation, Conceptualization, Funding acquisition, Data curation, Formal Analysis, Methodology, Visualization, Writing—original draft. RS: Writing—review and editing, Formal Analysis, Investigation, Methodology. JW: Investigation, Methodology, Writing—review and editing. RG: Writing—review and editing, Supervision. CU: Writing—review and editing, Formal Analysis. KC: Writing—review and editing, Conceptualization, Funding acquisition, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Science Foundation Research Traineeship Program (DGE-1828149) awarded to PI and the U.S. Department of Agriculture, Agricultural Research Service Project, 5050-21430-010-00D and USDA National Institute of Food and Agriculture AFRI (Award Number: 2017-67013-26212). Any opinions, findings, conclusion, or recommendation expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1330361/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

Bean plots of Fe, Zn in MI and NE, and Fe bioavailability in MI. The trials were carried out in 2018 and 2019 in MI and NE. Micronutrient traits were collected in one field replication, and two technical replications were done per sample. The mean of the two technical replications was used as a measure for each environment and year.

SUPPLEMENTARY FIGURE S2

Bean plots of yield (YD) and seed weight (SW) of the yellow panel. The trials were carried out in 2018 and 2019 in MI and NE. Best linear unbiased estimator (BLUEs) were obtained for each trait and trial, adjusting for spatial effects in the field.

SUPPLEMENTARY FIGURE S3

Principal Component Analysis of Andean Accessions from the YBC and Advanced Yellow Breeding Lines Using 18,357 SNPs. The corols represent the six parental YBC Accessions in Biparental Crosses.

SUPPLEMENTARY FIGURE S4

Genome-wide association analysis of Fe bioavailability (FeBio). The red line indicates the Bonferroni threshold at $\alpha = 0.05$. The black dashed lines highlight SNP-associated across years.

SUPPLEMENTARY FIGURE S5

Prediction accuracy of Fe and Zn concentration and seed yield (YD) in the 2019 prediction set comprised of 82 additional breeding lines. The models were trained using the Andean accession from the Yellow Bean Collection (YBC) and varying proportions of the prediction set (0%, 10% = 8, 20% = 16, 30% = 25) to train the models. The distribution of boxplots represents 100-fold cross-validations.

SUPPLEMENTARY FIGURE S6

Boxplots of Fe-Zn concentration and Fe bioavailability (FeBio) in the nine major seed types in the Yellow Bean Collection (YBC) in MI.

SUPPLEMENTARY FIGURE S7

Box plots of Fe, Zn, yield and seed weight (SW) of the yellow breeding lines categorized by family, evaluated in MI 2019.

SUPPLEMENTARY TABLE S1

Analysis of variance.

SUPPLEMENTARY TABLE S2

QTN identified by genome-wide association analysis using the FarmCPU model with a Bonferroni correction.

SUPPLEMENTARY TABLE S3

Prediction accuracy for micronutrient density and seed yield in the YBC using RKHS and sparse selection indices (SSI).

SUPPLEMENTARY TABLE S4

Prediction accuracy for Fe-Zn concentration and seed yield in the prediction set of 82 yellow breeding lines and the YBC using the RKHS and sparse selection indices (SSI).

SUPPLEMENTARY TABLE S5

Prediction accuracy for Fe-Zn concentration and seed yield in the prediction set of 82 yellow breeding lines and the Andean accession of the YBC using RKHS and sparse selection indices (SSI).

SUPPLEMENTARY TABLE S6

Shapiro-Wilk test results for phenotypic data in the Yellow Bean Collection.

References

- Ariza-Nieto, M., Blair, M. W., Welch, R. M., and Glahn, R. P. (2007). Screening of iron bioavailability patterns in eight bean (*Phaseolus vulgaris* L.) genotypes using the Caco-2 cell *in vitro* model. *J. Agric. Food Chem.* 55 (19), 7950–7956. doi:10.1021/jf070023y
- Beebe, S. (2020). Biofortification of common bean for higher iron concentration. *Front. Sustain. Food Syst.* 4 (November), 1–6. doi:10.3389/fsufs.2020.573449
- Bernardo, R. (2020). *Breeding for quantitative traits in plants*. Woodbury, MN, USA: Stemma Press.
- Blair, M. W., Astudillo, C., Rengifo, J., Beebe, S. E., and Graham, R. (2011). QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 122 (3), 511–521. doi:10.1007/s00122-010-1465-8
- Blair, M. W., and Izquierdo, P. (2012). Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor. Appl. Genet.* 125 (5), 1015–1031. doi:10.1007/s00122-012-1891-x
- Blair, M. W., Izquierdo, P., Astudillo, C., and Grusak, M. (2013). A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front. Plant Sci.* 4, 275–289. doi:10.3389/fpls.2013.00275
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103 (3), 338–348. doi:10.1016/j.ajhg.2018.07.015
- Caproni, L., Raggi, L., Talsma, E. F., Wenzl, P., and Negri, V. (2020). European landrace diversity for common bean biofortification: a genome-wide association study. *Sci. Rep.* 10 (1), 19775–19813. doi:10.1038/s41598-020-76417-3
- Cichy, K., Chiu, C., Isaacs, K., and Glahn, R. (2022). Dry bean biofortification with iron and zinc. *Biofortification Staple Crops*, 225–270. doi:10.1007/978-981-16-3280-8_10
- Cichy, K. A., Caldas, G. V., Snapp, S. S., and Blair, M. W. (2009). QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. *Crop Sci.* 49 (5), 1742–1750. doi:10.2135/cropsci2008.10.0605
- Cuevas, J., Crossa, J., Soberanis, V., Pérez-Elizalde, S., Pérez-Rodríguez, P., Campos, G. D. L., et al. (2016). Genomic prediction of genotype $\times$ environment interaction kernel regression models. *Plant Genome* 9 (3), 1–20. doi:10.3835/plantgenome2016.03.0024
- Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Zeffa, D. M., Nogueira, A. F., Ribeiro, L. A. B., et al. (2021). Genome-wide association study for grain mineral content in a Brazilian common bean diversity panel. *Theor. Appl. Genet.* 134 (9), 2795–2811. doi:10.1007/s00122-021-03859-2
- de los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A., and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res. (Camb.)* 92 (4), 295–308. doi:10.1017/S0016672310000285
- de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C., and Sorensen, D. (2013). Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* 9 (7), e1003608. doi:10.1371/journal.pgen.1003608
- FAO, IFAD, UNICEF, WFP, WHO (2022). "The State of Food Security and Nutrition in the World 2022. Repurposing food and agricultural policies to make healthy diets more affordable," in *Brief to the state of food security and nutrition in the world 2023* (Rome: FAO).
- Glahn, R. P., Lee, O. A., Yeung, A., Goldman, M. I., and Miller, D. D. (1998). Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an *in vitro* digestion/Caco-2 cell culture model. *J. Nutr.* 128 (9), 1555–1561. doi:10.1093/jn/128.9.1555
- Glahn, R. P., and Noh, H. (2021). Redefining bean iron biofortification: a review of the evidence for moving to a high Fe bioavailability approach. *Front. Sustain. Food Syst.* 5 (July), 1–9. doi:10.3389/fsufs.2021.682130
- Glahn, R. P., Wiesinger, J. A., and Lung'Aho, M. G. (2020). Iron concentrations in biofortified beans and nonbiofortified marketplace varieties in East Africa are similar. *J. Nutr.* 150 (11), 3013–3023. doi:10.1093/jn/nxaa193
- González-Camacho, J. M., de los Campos, G., Pérez, P., Gianola, D., Cairns, J. E., Mahuku, G., et al. (2012). Genome-enabled prediction of genetic values using radial

basis function neural networks. *Theor. Appl. Genet.* 125 (4), 759–771. doi:10.1007/s00122-012-1868-9

Gota, M., and Gianola, D. (2014). Kernel-based whole-genome prediction of complex traits: a review. *Front. Genet.* 5 (OCT), 1–13. doi:10.3389/fgene.2014.00363

Gunjača, J., Carović-Stanko, K., Lazarević, B., Vidak, M., Petek, M., Liber, Z., et al. (2021). Genome-wide association studies of mineral content in common bean. *Front. Plant Sci.* 12 (March), 636484–636512. doi:10.3389/fpls.2021.636484

Isidro, Y., Sánchez, J., and Akdemir, D. (2021). Training set optimization for sparse phenotyping in genomic selection: a conceptual overview. *Front. Plant Sci.* 12, 1–14. doi:10.3389/fpls.2021.715910

Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., and Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 131 (8), 1645–1658. doi:10.1007/s00122-018-3104-8

Izquierdo, P., Kelly, J. D., Beebe, S. E., and Cichy, K. (2023). Combination of meta-analysis of QTL and GWAS to uncover the genetic architecture of seed yield and seed yield components in common bean. *Plant Genome* 16, 203–288. e20420. doi:10.1002/tpg2.20328

Katuuramu, D. N., Hart, J. P., Porch, T. G., Grusak, M. A., Glahn, R. P., and Cichy, K. A. (2018). Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). *Mol. Breed.* 38 (4), 44. doi:10.1007/s11032-018-0798-x

Katuuramu, D. N., Wiesinger, J. A., Luyima, G. B., Nkalubo, S. T., Glahn, R. P., and Cichy, K. A. (2021). Investigation of genotype by environment interactions for seed zinc and iron concentration and iron bioavailability in common bean. *Front. Plant Sci.* 12 (May), 670965. doi:10.3389/fpls.2021.670965

Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., et al. (2020). Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front. Plant Sci.* 11, 1001. doi:10.3389/fpls.2020.01001

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. doi:10.1038/nmeth.1923

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27 (21), 2987–2993. doi:10.1093/bioinformatics/btr509

Lobaton, J. D., Miller, T., Gil, J., Ariza, D., Hoz, J. F., Soler, A., et al. (2018). Resequencing of common bean identifies regions of inter-gene pool introgression and provides comprehensive resources for molecular breeding. *Plant Genome* 11 (2), 170068. doi:10.3835/plantgenome2017.08.0068

Lopez-Cruz, M., Beyene, Y., Gowda, M., Crossa, J., Pérez-Rodríguez, P., and de los Campos, G. (2021). Multi-generation genomic prediction of maize yield using parametric and non-parametric sparse selection indices. *Hered. (Edinb.)* 127 (5), 423–432. doi:10.1038/s41437-021-00474-1

Lopez-Cruz, M., and de los Campos, G. (2021). Optimal breeding-value prediction using a sparse selection index. *Genetics* 218 (1), iyab030. doi:10.1093/genetics/iyab030

Lopez-Cruz, M., Dreisigacker, S., Crespo-Herrera, L., Bentley, A. R., Singh, R., Poland, J., et al. (2022). Sparse kernel models provide optimization of training set design for genomic prediction in multiyear wheat breeding data. *Plant Genome* 15 (July), e20254. doi:10.1002/tpg2.20254

Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., Mondal, S., et al. (2020). Regularized selection indices for breeding value prediction using hyper-spectral image data. *Sci. Rep.* 10 (1), 8195–8212. doi:10.1038/s41598-020-65011-2

Lorenz, A. J., and Smith, K. P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in Barley. *Crop Sci.* 55 (6), 2657–2667. doi:10.2135/cropsci2014.12.0827

MacDiarmid, J. I., and Whybrow, S. (2019). Nutrition from a climate change perspective. *Proc. Nutr. Soc.* 78 (3), 380–387. doi:10.1017/S0029665118002896

Mahajan, R., Zargar, S. M., Salgotra, R. K., Singh, R., Wani, A. A., Nazir, M., et al. (2017). Linkage disequilibrium based association mapping of micronutrients in common bean (*Phaseolus vulgaris* L.): a collection of Jammu and Kashmir. India. *Biotech* 7 (5), 295. doi:10.1007/s13205-017-0928-x

McClean, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., et al. (2018). White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytol.* 219 (3), 1112–1123. doi:10.1111/nph.15259

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Moraghan, J. T. (2004). Accumulation and within-seed distribution of iron in common bean and soybean. *Plant Soil* 264 (1–2), 287–297. doi:10.1023/b:plso.0000047762.32990.6b

Nazir, M., Mahajan, R., Mansoor, S., Rasool, S., Mir, R. A., Singh, R., et al. (2022). Identification of QTLs/candidate genes for seed mineral contents in common bean (*Phaseolus vulgaris* L.) through genotyping-by-sequencing. *Front. Genet.* 13 (March), 750814–750815. doi:10.3389/fgene.2022.750814

Perea, C., De La Hoz, J. F., Felipe Cruz, D., Lobaton, J. D., Izquierdo, P., Quintero, J. C., et al. (2016). Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics* 17 (5), 498. doi:10.1186/s12864-016-2827-7

Pérez-Rodríguez, P., and de Los Campos, G. (2022). Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics* 222 (1), iyac112. doi:10.1093/genetics/iyac112

Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., and Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat. Stat.* 23, 52–71. doi:10.1016/j.spasta.2017.10.003

Sadohara, R., Izquierdo, P., Couto Alves, F., Porch, T. G., Beaver, J. S., Urrea, C. A., et al. (2022). The Phaseolus vulgaris L. Yellow Bean Collection: genetic diversity and characterization for cooking time. *Genet. Resour. Crop Evol.* 69, 1627–1648. doi:10.1007/s10722-021-01323-0

Sadohara, R., Morris, D., Long, Y., Cichy, K., Izquierdo, P., and Urrea, C. A. (2021). Seed coat color genetics and genotype × environment effects in yellow beans via machine-learning and genome-wide association. *Plant Genome* 15, e20215. doi:10.1002/tpg2.20173

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). Supplemental info: a reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46 (7), 707–713. doi:10.1038/ng.3008

Smith, M. R., Golden, C. D., and Myers, S. S. (2017). Potential rise in iron deficiency due to future anthropogenic carbon dioxide emissions. *GeoHealth* 1 (6), 248–257. doi:10.1002/2016GH000018

StatGenGxE, V. B.-J. (2023). Genotype by environment (GxE) analysis. Available from: <https://cran.r-project.org/package=statgenGxE.2022>.

Tako, E., Reed, S., Anandaraman, A., Beebe, S. E., Hart, J. J., and Glahn, R. P. (2015). Studies of cream seeded carioca beans (*Phaseolus vulgaris* L.) from a Rwandan efficacy trial: *in vitro* and *in vivo* screening tools reflect human studies and predict beneficial results from iron Biofortified beans. *PLoS One* 10 (9), e0138479. doi:10.1371/journal.pone.0138479

Tello, D., Gil, J., Loaiza, C. D., Riascos, J. J., Cardozo, N., and Duitama, J. (2019). NGSEP3: accurate variant calling across species and sequencing protocols. *Bioinformatics* 35 (22), 4716–4723. doi:10.1093/bioinformatics/btz275

Uebesax, M. A., Cichy, K. A., Gomez, F. E., Porch, T. G., Heitholt, J., Osorno, J. M., et al. (2022). Dry beans (*Phaseolus vulgaris* L.) as a vital component of sustainable agriculture and food security—a review. *Legum. Sci.* 5 (April), 1–13. doi:10.1002/leg3.155

van Wonderen, D., Melse-Boonstra, A., and Gerdessen, J. C. (2023). Iron bioavailability should be considered when modeling omnivorous, vegetarian, and vegan diets. *J. Nutr.* 153 (7), 2125–2132. doi:10.1016/j.tjnut.2023.05.011

Voss-Fels, K. P., Cooper, M., and Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. *Theor. Appl. Genet.* 132 (3), 669–686. doi:10.1007/s00122-018-3270-8

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics Bioinforma.* 19 (4), 629–640. doi:10.1016/j.gpb.2021.08.005

Westfall, D., and Bauder, T. (2011). Zinc and iron deficiencies. *Color State Univ. Ext.* 545.

Wiesinger, J. A., Cichy, K. A., Glahn, R. P., Grusak, M. A., Brick, M. A., Thompson, H. J., et al. (2016). Demonstrating a nutritional advantage to the fast-cooking dry bean (*Phaseolus vulgaris* L.). *J. Agric. Food Chem.* 64 (45), 8592–8603. doi:10.1021/acs.jafc.6b03100

Wiesinger, J. A., Cichy, K. A., Tako, E., and Glahn, R. P. (2018). The fast cooking and enhanced iron bioavailability properties of the manteca yellow bean (*Phaseolus vulgaris* L.). *Nutrients* 10 (11), 1609. doi:10.3390/nu10111609

Wiesinger, J. A., Glahn, R. P., Cichy, K. A., Kolba, N., Hart, J. J., and Tako, E. (2019). An *in vivo* (Gallus gallus) feeding trial demonstrating the enhanced iron bioavailability properties of the fast cooking manteca yellow bean (*Phaseolus vulgaris* L.). *Nutrients* 11 (8), 1768. doi:10.3390/nu11081768

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., et al. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *Lancet* 393 (10170), 447–492. doi:10.1016/S0140-6736(18)31788-4