Marker–Trait Association Analysis of Iron and Zinc Concentration in Lentil (Lens culinaris Medik.) Seeds

Hamid Khazaei, Rajib Podder, Carolyn T. Caron, Shudhangshu S. Kundu, Marwan Diapari, Albert Vandenberg, and Kirstin E. Bett*

Abstract

Lentil (Lens culinaris Medik.) seeds are relatively rich in iron (Fe) and zinc (Zn), making lentil a potential crop to aid in the global battle against human micronutrient deficiency. Understanding the genetic basis for uptake of seed Fe and Zn is required to increase sustainable concentrations of these minerals in seeds. The objectives of this study were to characterize genetic variation in seed Fe and Zn concentration and to identify molecular markers associated with these traits across diverse lentil accessions. A set of 138 cultivated lentil accessions from 34 countries were evaluated in four environments (2 sites × 2 yr) in Saskatchewan, Canada. The collection was genotyped using 1,150 single-nucleotide polymorphism (SNP) markers that are distributed across the lentil genome. The germplasm tested exhibited a wide range of variation for seed Fe and Zn concentration. The marker–trait association analysis detected two SNP markers tightly linked to seed Fe and one linked to seed Zn concentration (−log10 P ≥ 4.36). Additional markers were detected at −log10 P ≥ 3.06. A number of putative candidate genes underlying detected loci encode Fe- and Zn-related functions. This study provides insight into the genetics of seed Fe and Zn concentration of lentil and opportunities for marker-assisted selection to improve micronutrient concentration as part of micronutrient biofortification programs.

Core Ideas

- First study to provide knowledge on genetics of biofortification in lentils.
- SNP markers associated with seed Fe and Zn concentration were identified in lentils.
- MAS to enhance breeding programs aiming to fight global micronutrient malnutrition.

Lentil (Lens culinaris Medik.) is an annual, self-pollinated, highly nutritious grain legume crop (pulse) with a haploid genome size of approximately 4 Gbp (2n = 2x = 14). Globally, lentil is the fifth most important grain legume after soybean (Glycine max (L.) Merr.), common bean (Phaseolus vulgaris L.), chickpea (Cicer arietinum L.), and pea (Pisum sativum L.). Global lentil production has risen about 350% in the past 40 yr (FAO, 2015), which is in sharp contrast to other pulse crops. Grain legume seeds are typically rich in protein, fiber, complex carbohydrates, and essential micronutrients such as iron (Fe) and zinc (Zn). Micronutrient deficiency is recognized as the “hidden hunger” in the world. Both Fe and Zn are essential for sustenance and optimal physiological function of all life forms on the planet (Bailey et al., 2015). Nearly...
one-third and one-fifth of the world’s population are Fe- and Zn deficient, respectively (de Benoist et al., 2008). Lentil seeds have excellent Fe and Zn concentration compared with many grain legumes and cereals (Hemalatha et al., 2007; Ray et al., 2014; Sahuquillo et al., 2003; Thavarajah et al., 2011), which is contributing to lentil gaining in importance as a staple food to combat human micronutrient malnutrition globally.

Genetic diversity for seed Fe and Zn concentration has been investigated in lentil for improvement through conventional breeding (Karaköy et al., 2012; Kumar et al., 2014, 2016; Singh et al., 2017). However, identification of molecular markers for selecting genes associated with increased micronutrient accumulation is just starting. For instance, four quantitative trait loci regions and 36 putative molecular markers controlling genes for selenium (Se) uptake in lentil seeds were identified using a biparental population developed from the cross PI 320937 × ‘Eston’ (Ates et al., 2016).

In contrast to biparental population-based linkage mapping, association mapping (AM) allows for exploitation of natural diversity across a wider array of genotypes. It can be used to assess the historical and evolutionary recombination events that occurred at the population level, resulting in detection of additional regions or even candidate genes associated with the trait (Nordborg and Tavaré, 2002; Zhu et al., 2008). Multiple levels of recombination have accumulated in landraces, natural populations, breeding lines, and commercial varieties, which increases the resolution of association mapping detection. Association mapping is based on linkage disequilibrium (LD), which is the nonrandom associations of alleles in haplotypes of a population. The AM method has been used increasingly over the last decade for the study of complex genetic traits in many plant species. A limited number of studies have reported marker-micronutrient (Fe, Zn, and Se) associations in pulse crops such as chickpea (Diapari et al., 2014; Upadhyaya et al., 2016) and field pea (Cheng et al., 2015; Diapari et al., 2015; Kwon et al., 2012). A few studies have been conducted on the genetics of micronutrient accumulation in seeds of several cereals using the AM approach, for example in barley (Hordeum vulgare L.) (Mamo et al., 2014), rice (Oryza sativa L.) (Norton et al., 2014; Nawaz et al., 2015; Huang et al., 2015), and a synthetic hexaploid wheat (Gorafi et al., 2016). To date there is no information on the genomic regions controlling seed Fe and Zn accumulation in lentil seeds that could be deployed in marker-assisted selection. Thus, the current study was designed to determine the genetic basis of seed Fe and Zn concentration in lentil by using AM to determine marker–trait associations using a single-nucleotide polymorphism (SNP) array derived from cultivated lentil sequences.

Materials and Methods

Plant Material

A total of 138 cultivated lentil accessions, originating from 38 countries across the three major agro-ecological zones of lentil production (Khazaei et al., 2016), were collected from various sources, including the Crop Development Centre collection in Saskatoon, Canada; the International Center for Agricultural Research in the Dry Areas; and the USDA–ARS. The accessions and their origins are shown in Supplemental Table S1A.

Experimental Conditions

Experiments were conducted at two different locations: one at the Saskatchewan Pulse Growers land (SPG, 52.05° N, 106.61° W) and one at the University of Saskatchewan experimental farm in Sutherland (52.14° N, 106.61° W) in 2013 and 2014. The experiment was set up in a randomized complete block design, with six and three replicates in 2013 and 2014, respectively. Sowing date was in the last week of May in both years. The average rainfall during the growing cycle (May–August) was 19% higher in 2014 (218.9 mm) than in 2013 (184.1 mm) (Supplemental Table S2). Experimental plots consisted of around 60 seeds of each genotype sown in three-row, 1.0-m² plots with 0.25 m distance between rows. The soil at these locations was Brown and Dark Brown Chernozem. The soil at the Sutherland farm had greater amounts of extractable Fe than the soil at SPG; however, no significant differences for Zn were present between the two locations (Supplemental Table S3).

Phenotyping

Each plot was harvested separately at maturity. Plot samples were threshed, and seed samples (~5 g) were carefully washed with distilled water prior to microanalytical analysis. Seeds were ground to a fine powder using a Cyclone Sample mill (Model 3010–030, UDY). A 0.5-g ground sample was digested in a 30 mL digestion tube following a HNO₃–H₂O₂ digestion method (Alcock, 1987) using an automatic Vulcan 84 digester (Questron Technology). All chemicals used for digestion were analytical grade. In the digestion chamber, samples were digested with 6 mL of HNO₃ at 90°C for 45 min followed by an addition of a 5 mL 30% H₂O₂ and continued digestion for another 65 min. Then, the solutions were reduced with 3 mL of HCl (6 M), and digestion continued for another 5 min before the sample solutions were cooled down. All the sample solutions were then diluted with deionized water (18 MΩ) and adjusted to a volume of 25 mL. The Fe and Zn concentrations were analyzed using a flame atomic absorption spectrophotometer (Nova 300, Analytic Jena AG), which was calibrated using standard solutions containing different amount of Zn and Fe. Iron and zinc were analyzed on the basis of absorption of optical radiation at 248.3 and 213.9 nm wavelengths, respectively. The National Institute of Standards and Technology standard reference material 1573a (tomato
leaves, Fe = 368 ± 7 mg kg⁻¹, Zn = 30.9 ± 0.5 mg kg⁻¹) was used for evaluation of the method.

Statistical Analysis
We used SAS v. 9.4 for micronutrient data analysis. The datasets from the germplasm phenotyping for Fe and Zn were subjected to a mixed-model ANOVA (SAS Institute, 2015). The two locations were combined to form location-genotypes. The seed Fe and Zn concentrations were considered dependent variables. Among the other four variables, replications were considered as random effects, and genotypes, year, and location were considered as fixed effects. Broad-sense heritability (H²) of the Fe and Zn concentrations was calculated as the ratio of genotypic variance (σ²g) to phenotypic variance (σ²p) using SAS v. 9.4.

Genotyping
All samples were genotyped with the Lc1536 GoldenGate array as described previously by Sharpe et al. (2013). All sequence and marker information is accessible via the KnowPulse web portal (http://knowpulse.usask.ca/portal/project/Lc1536-Golden-Gate-Assay). Of the 1243 polymorphic SNP markers, 1150 markers were used for association mapping analysis after markers with minor allele frequency <0.05 were excluded.

Population Structure and Genetic Diversity
STRUCTURE v. 2.3.4 (Pritchard et al., 2000) was used to calculate the most probable number of subpopulations (K). Five independent runs were done for each K ranging from 1 to 10 with both a burn-in time and Markov Chain Monte Carlo replication number of 500,000. Selection of the best K value was based on the procedure presented in Evanno et al. (2005) by submitting the results for each K to the STRUCTURE HARVESTER website, which returned the L(K) and ΔK value (Earl and vonHoldt, 2012). A Q-matrix was obtained from the membership probability of each individual and was used for further association mapping.

The marker-based relative kinship analysis estimates the approximate identity between two given individuals over the average probability of identity between two random individuals. Relative kinship provides useful inheritance information. The kinship matrix of the 138 accessions was obtained using the method by Loiselle et al. (1995) with SPAGeDi software (Hardy and Vekemans, 2002). The relative kinship matrix (K) represents the degree of identity between two given individuals (accessions). Negative values between two individuals, indicating little or no relationship, were altered to “0.”

Association Analysis
TASSEL (Trait Analysis by Association, Evolution, and Linkage) v. 5.2.31 (Bradbury et al., 2007) was used to perform association mapping analysis. Associations between SNP markers and phenotypes (seed Fe and Zn concentration) were evaluated using a mixed linear model (MLM) by incorporating genotypes, phenotypes, principal component analysis (PCA), population structure (Q), and kinship matrix (K). Both MLM (PCA + K) and MLM (Q + K) models were used. The PCA matrix was calculated using TASSEL. To report the most robust marker–trait associations, we used Bonferroni-corrected thresholds at α = 1, α = 0.5, and α = 0.05 as the cut-offs (8.69 × 10⁻⁴, 4.35 × 10⁻⁵, and 4.35 × 10⁻⁶, corresponding to “− log10 P (a/n)” values of 3.06, 3.36, and 4.36, respectively) to account for multiple testing (Yang et al., 2013).

Results
Phenotypic Variation
The 138 accessions differed for seed Fe and Zn concentrations (P < 0.001). In both years, variability in seed Fe and Zn concentrations was higher from the SPG location than for the Sutherland location. There was a nearly twofold difference in range for both traits in all site-year combinations. The average seed Fe and Zn concentrations were 76.2 ± 9.5 and 37.8 ± 3.5 ppm, respectively (Table 1). A significant genotype × year interaction was observed only for Fe concentration. A significant year effect was also noted for both Fe and Zn. The broad-sense heritability was higher for seed concentration of Fe compared with that of Zn in both years (Table 2).

Relative Kinship
About 72% of the pairwise kinship estimates were zero or close to zero, suggesting that these accessions are unrelated. The remaining estimates ranged from 0.1 to 1.2 (Supplemental Fig. S1; Supplemental Table S1B), with an exponentially decreasing number of pairs falling into higher estimate categories. The kinship analysis indicated complex familial relationships among the 138 accessions, which complements the known pedigree history.

Population Structure and Genetic Relationship
The estimated log probability of the data [LnP(K)] for each K plateaued at K = 2. The maximum ΔK value was also reached at K = 2. The majority of accessions in one of the groups were Canadian breeding lines and cultivars. The second group had a mixture of ICARDA breeding lines and accessions from the collection held by the USDA genebank (Supplemental Table S1A).

Marker-Trait Associations
Both MLM (Q + K) and MLM (PCA + K) association tests of seed Fe and Zn concentration identified several significantly linked SNP markers (Fe concentration, nine markers; Zn concentration, 12 markers) with a significance level of at least − log10 P ≥ 3.06. Two of these SNP markers had a strong association (− log10 P ≥ 4.36) with seed Fe concentration and one with seed Zn concentration (Fig. 1). Significant SNP markers found in two or more datasets (site and year) are regarded as more reliable than those significant only in a single dataset. Most significant markers revealed associations across multiple environments (sites and years) as well as with two different MLM methods (Q and PCA). LcC25737p350 and LcC24316p626
had strong association with seed Fe concentration at both SPG and Sutherland across 2013 and 2014. LcC06625p437 was highly significant for Fe in 2014 only (Fig. 1).

Overall, weaker associations were detected for seed Zn concentration than for seed Fe concentration. Most of the significant markers for Zn were detected in 2013 but were absent in 2014. Although LcC06739p564 (\(-\log_{10} P \geq 3.36\)) and, to a lesser extent, LcC04105p1090 (\(-\log_{10} P \geq 3.06\)) and LcC03907p461 (\(-\log_{10} P \geq 3.36\)) were the most common markers across sites and statistical models, LcC05435p444 showed the strongest association (\(-\log_{10} P \geq 4.36\)) with Zn concentration based on the PCA method (Fig. 1).

In general, the percentage of phenotypic variation explained by each marker (\(R^2\)) ranged from 9 to 21% for the two traits. The results also indicated that five out of nine SNP markers associated to seed Fe concentration are located on lentil chromosome 5, although no preponderance on any particular chromosome was evident for associations with seed Zn concentration.

The genes associated with the SNP markers identified as significant for Zn or Fe and their location on the L. culinaris genome (version 1.2) are provided in Supplemental Table S1C. The region surrounding these markers can be explored using the lentil JBrowse at http://knowpulse.usask.ca/portal/jbrowse/Lentil.

**Discussion**

Association mapping is a powerful tool for identifying alleles and loci responsible for economically important traits, especially due to the availability of affordable high-throughput genotyping technologies in crop species. This work presents the first association mapping analysis in lentil that detects genetic markers associated with seed micronutrient concentration (Fe and Zn). The germplasm surveyed here was representative of that grown in the major lentil production areas in the world and should capture the range of genotypes currently in use globally. Our results identified associations between SNP markers and seed Fe and Zn concentration, revealing the potential to apply these detected molecular markers for marker-assisted selection in lentil to improve the seed micronutrient concentration.

The results of this research revealed that there is considerable variation for Fe and Zn concentration in seeds of the lentil germplasm surveyed (Table 1). This should allow breeding programs to exploit the available diversity for improvement of seed Fe and Zn biofortification in this staple food crop. On average, the Fe concentration was slightly (7%) higher in seeds of Canadian breeding lines (79.6 ± 6.2 ppm) compared with those in international germplasm (74.1 ± 10.5 ppm), as also noted by Thavarajah et al. (2011), although greater variation existed within the international collection. The extremes of seed Fe concentration were found within the imported germplasm (minimum: ILL 2607 from India; maximum: PI 299215 from Chile) (Supplemental Table S1D). Less variation was observed within the entire germplasm set for seed Zn concentration compared with that for Fe. The lowest seed Zn concentration was observed in ILL 2194 from Pakistan, whereas ILL 4875 from Uzbekistan had highest (Supplemental Table S1D). A strong positive correlation was observed between seed Fe and Zn concentrations (Fig. 2), a trend that has been reported in several different crop species (e.g., barley [Mamo et al., 2014], pea and chickpea...
This may suggest a common genetic mechanism responsible for accumulation and final concentration of seed Fe and Zn, as described for chickpea (Diapari et al., 2014), common bean (Blair et al., 2009), maize (Qin et al., 2012), and rice (Anuradha et al., 2012). However, no common molecular markers were associated with both seed Fe and Zn concentration, suggesting that improvement of seed concentration may require independent genetic selection in lentil.

The level of LD as an indicator of power/resolution of association mapping to identify the large effect association with Fe and Zn was estimated by Fedoruk (2013) for the current SNP marker set over the same germplasm set. On average, LD was relatively low ($r^2 = 0.1$ within map distance of 5 cM). Selfing, as well as bottlenecks imposed by selecting for adaptation to a particular agro-ecological region in this species, should lead to higher LD due to low effective recombination rates (reviewed by Huang and Han [2014]). Overall, this should mean that there are

---

**Fig. 1.** Single-nucleotide polymorphism markers associated with Fe and Zn concentrations in seeds of lentil using mixed linear model (population structure (Q) and principal component analysis (PCA)) models in Saskatchewan Pulse Growers (SPG) and Sutherland (STH), Saskatchewan, Canada, in 2013 and 2014. Red, $P \leq 4.35 \times 10^{-5}$; orange, $P \leq 4.35 \times 10^{-4}$; yellow, $P \leq 8.69 \times 10^{-4}$; gray (weak), $P = 0.001$. $R^2$ values are in order of each association from left to right. K, kinship matrix; ND, not determined.

---

**Fig. 2.** Relationship between lentil seed Fe and Zn concentration ($n = 138$).
reasonable levels of LD for effective association mapping with the available SNP marker density. A total of three SNP markers were found closely associated with seed Fe or Zn concentrations. We found fewer associations overall as compared with studies in other pulses (Diapari et al., 2014, 2015; Upadhyaya et al., 2016). This might be due to lower levels of genetic variability for these traits in the screened germplasm or to the larger size of the lentil genome requiring a greater number of molecular markers. The importance of those two factors has been revealed in previous association studies (e.g., Myles et al., 2009; Racedo et al., 2016; Zhu et al., 2008). The results also suggest that loci controlling Fe concentration are most likely located in chromosome 5 of lentil, although the markers detected for seed Zn concentration were spread throughout the lentil genome.

Population structure is a confounding concern in marker–trait association studies because it can elevate the likelihood of false-positive associations. A Q matrix generated by the program STRUCTURE presents an overall picture of population disparity (Lander and Schork, 1994; Zhao et al., 2007), whereas a PCA summarizes relatedness in genome patterns (Price et al., 2006). Moreover, running STRUCTURE is computationally intensive and perhaps not feasible on very large datasets, especially because the size of datasets continues to increase with new sequencing technologies. Using innovative methods such as ADMIXTURE or sparse non-negative matrix factorization to find the optimal number of subpopulations (K) is also considered computationally intensive (Popescu et al., 2014). It is also known that fewer residual false-positive associations occur when using the PCA model (Yang et al., 2011). In our study, we used both PCA + K and Q + K models to identify markers associated with Fe and Zn concentration and in an effort to minimize both false-positive and false-negative associations. Our results showed only slight differences between the outcomes of the Q + K and PCA + K models. In fact, all highly significant associations were discovered using the PCA model (Fig. 1). The same trend has been reported in Arabidopsis (Zhao et al., 2007) and rapeseed (Brassica napus L.) (Li et al., 2014; Wang et al., 2016). These results suggest that the computationally simple PCA model can be substituted for STRUCTURE-generated Bayesian clustering analysis (Q matrix) in the MLM approach.

Iron and Zn taken up by the root are transferred cell to cell with metal transporters and are ultimately allocated to the seed cotyledon (Grusak, 2002). Biofortification pathways for Fe (Conte and Walker, 2011; Waters and Sankaran, 2011) and Zn (White and Broadley, 2011) accumulation have been extensively studied in model plant species. A list of the genes containing the SNP markers (Sharpe et al., 2013) and their location on the lentil genome can be found in Supplemental Table S1C. While expecting to find a marker in a candidate gene when only surveying 1150 of the expected >30,000 genes in the genome is tenuous, it is worth looking for candidate genes at the marker and in the region up and downstream of it. Doing so in this case revealed some potential candidates for controlling the observed variation. For example, the sequence of marker LcC06625p437 was found within a cytosolic Fe-S cluster assembly factor NUBP1-like protein in a well-conserved region of Mt chromosome 1 and also in Arabidopsis (Bernard et al., 2013). The possible function of this protein in plant cells is to coordinate Fe uptake, storage, and bioavailability. Additionally, genes for transmembrane proteins and transporters were found in the marker regions. Transmembrane proteins are known to play important roles in Fe uptake from the root and cell–cell transportation of Fe in plants (Dubeaux et al., 2015). Zinc uptake is less well defined, but in Arabidopsis it is likely performed by Zn transporters of the ZIP family (Waters and Sankaran, 2011). They are a group of regulatory proteins that have diverse functions in plant species, including abiotic stress adaptation and Zn transportation (Takatsuji, 1999). These findings could be used to accelerate our understanding of genes/pathway mechanisms related to micronutrients uptake in grain legume seeds.

Conclusions

This study provides insight into the genetic basis of variability in seed Fe and Zn concentration in a diverse set of lentil and provides valuable SNP markers that could be used in conjunction with the high micronutrient accessions to breed for increased levels of seed Fe and Zn in lentil. The results of this study confirmed the abundance of seed Fe in lentil too. We aim to validate these markers on a broad-range of wild and cultivated lentil populations. The selected marker then will be used for marker-assisted selection to improve Fe and Zn concentration in lentil seeds with desirable end-user qualities.

Acknowledgments

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and the Saskatchewan Pulse Growers. The authors thank Barry Goetz, Rob Stonehouse, Brent Barlow, and Lacey-Anne Sanderson for technical support of the experiments.

References


