Common bean (*Phaseolus vulgaris* L.) with increased cysteine and methionine concentration

Rosa Cecilia Viscarra-Torrico1,2 | Aga Pajak1 | Alvaro Soler Garzón3 | BaiLing Zhang4 | Sudhakar Pandurangan1 | Marwan Diapari1 | Qijian Song5 | Robert L. Conner6 | James D. House7 | Phillip N. Miklas8 | Anfu Hou6 | Frédéric Marsolais1,2

1London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario, Canada
2Department of Biology, University of Western Ontario, London, Ontario, Canada
3Irrigated Agriculture Research and Extension Center, Washington State University, Prosser, Washington, USA
4Harrow Research and Development Centre, Agriculture and Agri-Food Canada, Harrow, Ontario, Canada
5Beltsville Agricultural Research Center, USDA-ARS, Beltsville, Maryland, USA
6Morden Research and Development Centre, Agriculture and Agri-Food Canada, Morden, Manitoba, Canada
7Department of Food and Human Nutritional Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
8Grain Legume Genetics and Physiology Research Unit, USDA-ARS, Prosser, Washington, USA

Abstract

In common bean (*Phaseolus vulgaris* L.), seed storage protein deficiency is associated with increased total cysteine and methionine concentration. The goal of this study was to generate germplasm lines that combine this characteristic with adaptation to short season conditions in Manitoba, Canada. A recombinant inbred line population was developed by crossing the storage protein deficient genotype, SMARC1N-PN1 with the cultivar Morden003. Two lines, 2-37 and 3-84, with a stable protein profile over 2 years at two locations were identified. Like SMARC1N-PN1, both lines had a significantly higher cysteine concentration than Morden003, by approximately 35%. Methionine levels were elevated by approximately 15%, while tryptophan levels were also increased by approximately 30%. Line 2-37 had a significant increase in protein quality, as measured by in vitro protein digestibility corrected amino acid score, by approximately 40%, as compared with Morden003. The increased protein quality for this line is attributable to higher levels of total cysteine and methionine, while having an overall reduction in crude protein concentration. Line 2-37 had a similar seed yield as SMARC1N-PN1, with a maturity comparable to Morden003. The results of high-density single nucleotide polymorphism (SNP) genotyping and quantitative trait locus analysis of recombinant inbred lines indicated that variation in cysteine concentration was determined by the phaseolin locus, while variation in methionine concentration was determined by both the phaseolin and lectin loci. SNP markers that track the introgression of phaseolin and lectin deficiency into the Morden003 background were identified and validated.

KEYWORDS

navy bean, phaseolin, protein quality, seed storage protein, sulfur containing amino acids, tryptophan, whole protein
1 | INTRODUCTION

Common bean (dry bean, *Phaseolus vulgaris*) plays a major role in addressing the nutritional needs of a growing increasing human population, with its importance as a crop in Sub-Saharan Africa and South and Central America. Common bean constitutes primarily a source of protein in human diets (De Ron et al., 2015). Like other pulses, the quality of its protein, as defined by the balanced composition of nutritionally essential amino acids, is limited first and foremost by the low levels of sulfur amino acids, methionine and cysteine, and to some extent tryptophan (Nosworthy et al., 2017). In combination with cereals, common bean, like other legumes, can provide a balanced source of protein. Nevertheless, from a biofortification standpoint, improving protein quality of common bean represents a valuable goal. A recent example from cereal crops is the development of improved Quality Protein Maize, with enhanced levels of both lysine and methionine (Planta & Messing, 2017). Improving protein quality of common bean and other pulses is also relevant to nutritional claims on protein content (Wiggins et al., 2018).

Globulins are a group of seed storage proteins being part of the cupin superfamily. The 7S globulin, phaseolin, normally constitutes approximately half of total seed protein in cultivated varieties of common bean (Vitale & Bollini, 1995). Phaseolins are encoded by a single, complex locus on chromosome 7 containing multiple genes in tandem (Joshi et al., 2017; Pandurangan & Marsolais, 2017; Talbot et al., 1984). Despite the fact that phaseolins have few methionine residues, it was previously reported that the total concentration of methionine is positively correlated with phaseolin levels, since they are so abundant in seed (Gepts & Bliss, 1984). Montoya et al. (2010) proposed to use phaseolin isoforms from wild accessions as a possible strategy to improve protein quality. Lectins are the second most abundant seed proteins after phaseolins and account for approximately 5%–10% of total seed protein. Lectins protect seeds from insect pests and herbivory. Most lectins are encoded at a single, complex arcelin/phytohemagglutinin/α-amylase inhibitor (APA) locus on chromosome 4 (Freyre et al., 1998; Osborn et al., 1986).

We previously reported that a progressive deficiency in major seed proteins, phaseolin and lectins, results in a significant increase in sulfur amino acid concentration, in a series of genetically related lines, with cysteine concentration elevated by up to 70% and methionine concentration by 10%–20% (Taylor et al., 2008). The changes in sulfur amino acid levels happen through proteome rebalancing, to compensate for the absence of phaseolin and lectins, which favors the accumulation of sulfur-rich proteins (Liao et al., 2012; Marsolais et al., 2010; Pandurangan et al., 2016; Pandurangan, Sandercock, et al., 2015; Yin et al., 2011). In the present work, we hypothesized that the characteristics identified in SMARC1N-PN1 could be transferred to other germplasm lines in a different genetic background. SMARC1N-PN1, deficient in phaseolin and major lectins (Osborn et al., 2003), was crossed with the cultivar Morden003 (Mündel et al., 2004), to recover lines having increased cysteine and methionine concentration, with adaptation to short season conditions in Manitoba, Canada. High density single nucleotide polymorphism (SNP) genotyping and quantitative trait loci (QTL) mapping was performed to determine the effect of genetic loci on sulfur amino acid profiles.

2 | MATERIALS AND METHODS

2.1 | Plant material

A total of 183 $F_{2:8}$ recombinant inbred lines (RILs) were developed through a single-seed descent method from a cross between SMARC1N-PN1 and Morden003. “Morden003” is a navy bean cultivar with early maturity and primary adaptation to southern Manitoba, Canada (Mündel et al., 2004). SMARC1N-PN1 is a navy bean germplasm line in ‘Sanilac’ genetic background, with deficiency in phaseolin and lectin polypeptides (Osborn et al., 2003). The initial reciprocal $F_1$ cross-pollination was made in 2008 in the greenhouse at the Morden Research and Development Centre, Morden, MB. The $F_1$ plants were grown in the greenhouse and three true hybrid plants were identified with the growth morphology resembling both parental lines. Single seed descent was adopted from the individual $F_2$ seed in the greenhouse at Morden Research and Development Centre until $F_{2:8}$ generation. The RILs derived from the three individual $F_1$ plants were designated as Pop1, Pop2, and Pop3. Every single plant was bulk-harvested at the last generation in 2013. To make sure that pure seeds were available for future analysis, an additional single seed descent increase of the RILs was performed in 2014 ($F_{2:9}$).

2.2 | Field experiment conditions

In 2014 and 2015, the 183 lines were grown in three separate yield trials of Pop1 (66 lines), Pop2 (53 lines), and Pop3 (64 lines) along with the two parental lines for each trial at Morden Research and Development Centre (49.1923°N, 98.0977°W, Elevation 297 m) and at London Research and Development Centre, London, ON, Canada (43.0304°N, 81.2087°W, Elevation 271 m). For both locations, fertilizers were applied to meet the plant growth requirement levels of N (112 kg ha$^{-1}$) and P (50 kg ha$^{-1}$). For only the Morden Research and Development Centre trial, S (22 kg ha$^{-1}$) was also added each year. Each trial consisted of two rows per line with 5 m length and 75 cm row spacing, in a randomized complete block design with three replications. Each plot was harvested at natural (no dessicants) harvest maturity in bulk. The average growing season minimum and maximum temperatures in Morden were 10.8°C and 11.7°C, and 23.6°C and 25.7°C, respectively, during the 2 years of the study. The average growing season minimum and maximum temperatures in London were 12.2°C and 13.2°C, and 24.8°C and 24.6°C, respectively. The total precipitation for 2014 and 2015, respectively, during the growing seasons was 243.9 mm and 171.5 mm in Morden, and 249.7 mm and 269.5 mm in London.
2.3 | Determination of phaseolin and lectin protein profiles

Seed samples (1.2 g) from the F_{2:9} generation collected in the greenhouse were ground with a Kleco ball mill (Garcia Machine, Visalia, CA). Soluble protein was extracted from 50 mg of ground tissue using 1 ml of cold extraction buffer (50 mM Tris-HCl pH 8.0, 50 mM KCl, 1 mM CaCl_2, 10% glycerol) containing 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM dithiothreitol (DTT). Extracts were centrifuged twice at 25,000 × g for 20 min at 4°C. A total of 4.5 μl of the 4 × SDS protein sample buffer was added to 1.5 μl of extract. Samples were boiled for 5 min at 99°C. Five μl of the protein sample was loaded on an Invitrogen™ NuPAGE™ Novex 4%-12% Bis-Tris Protein Gel (https://www.thermofisher.com/ca/en/home/brands/invitrogen.html), run with NuPAGE™ 1 × MES SDS Running Buffer and visualized by Coomassie staining.

2.4 | Amino acid analysis

Seed samples from the F_{2:9} generation collected in the greenhouse were prepped and total amino acids analyzed as described by Jafari et al. (2016). Samples were acid hydrolysed in the presence of 6 N HCl and 1% phenol. Hydrolysis was performed using an Eldex Workstation (Napa, CA). NorValine was used as an internal standard. Cysteine was detected separately as cysteic acid, after performic acid oxidation. The sample was dried under vacuum for 45 minutes, resuspended in HPLC water and filtered using Nanosep MF 0.2 μm centrifugal devices (Pall, Mississauga, ON). The sample was derivatized with O-phthalaldehyde and 3-mercaptopropionic acid as per the manufacturer’s instructions (Agilent Technologies, Mississauga, ON). Amino acid analysis was performed on an Agilent 1260 Infinity HPLC system. Data were collected, stored and processed using OpenLab CDS software. Amino acids were detected on the basis of retention time established for standards. Linearity of the peak areas at different concentrations was determined. Calculations were based on the area under the peak for a known concentration. For a subgroup of lines, amino acids were analyzed by HPLC and 3-mercaptopropionic acid as per the manufacturer’s instructions (Agilent Technologies, Mississauga, ON). Amino acid analysis was performed on an Agilent 1260 Infinity HPLC system. Data were collected, stored and processed using OpenLab CDS software. Amino acids were detected on the basis of retention time established for standards. Linearity of the peak areas at different concentrations was determined. Calculations were based on the area under the peak for a known concentration. For a subgroup of lines, amino acids were analyzed by HPLC and 3-mercaptopropionic acid as per the manufacturer’s instructions (Agilent Technologies, Mississauga, ON). Amino acid analysis was performed on an Agilent 1260 Infinity HPLC system. Data were collected, stored and processed using OpenLab CDS software. Amino acids were detected on the basis of retention time established for standards. Linearity of the peak areas at different concentrations was determined. Calculations were based on the area under the peak for a known concentration. For a subgroup of lines, amino acids were analyzed by HPLC and 3-mercaptopropionic acid as per the manufacturer’s instructions (Agilent Technologies, Mississauga, ON). Amino acid analysis was performed on an Agilent 1260 Infinity HPLC system. Data were collected, stored and processed using OpenLab CDS software. Amino acids were detected on the basis of retention time established for standards. Linearity of the peak areas at different concentrations was determined. Calculations were based on the area under the peak for a known concentration.

2.5 | Measurement of crude protein and protein digestibility

Seed nitrogen concentration was measured at the Lethbridge Research and Development Centre of Agriculture and Agri-food Canada using a Flash 2000 Elemental Analyzer (Thermo Fisher Scientific). Crude protein is expressed as nitrogen concentration × 6.25 on a dry weight basis. Seed moisture was measured with a Grain Analysis Computer (DICKEY-john, Auburn, IL). Protein digestibility was determined as follows. Seeds were boiled for 1 hr in ultrapure water, drained and freeze dried for 16 hr. Dried seeds were ground to a fine powder using a Kleco ball mill. The multi-enzyme solution for in vitro protein digestibility was prepared by mixing 31 mg of chymotrypsin (bovine pancreas, ≥40 units/mg protein), 16 mg of trypsin [porcine pancreas, 13,000-20,000 N-α-benzoyl-L-arginine ethyl ester (BAEE) units/mg protein] and 13 mg of protease (Streptomyces griseus, ≥15 units/mg solid) in 10 ml of ultrapure water. The pH of the solution was adjusted to 8.0 using 0.1 M NaOH and HCl and the solution was stored at 37°C. Two hundred milligrams of flour was added to 10 ml of ultrapure water. The solution was stirred for 1 hr at 37°C and pH adjusted to 8.0 before adding 1 ml of the multi-enzyme solution. The pH of the protein solution was monitored and recorded every 1 min for 10 min and the in vitro protein digestibility (IVPD) was calculated as follows:

$$\text{IVPD} = 65.66 + 18.10 \times \Delta \text{pH}_{10\text{min}}$$

where $\Delta \text{pH}_{10\text{min}}$ refers to the change in pH from the initial value of 8.0 to the end of the 10-min period.

2.6 | High density genotyping

For genotyping using the Infinium II BARCBean6K_3 BeadChip (Illumina, Vancouver, BC), DNA was isolated as described in Pandurangan et al. (2016) using the DNeasy Plant Mini Kit (Qiagen, Toronto, ON). Hybridization was carried out as described by Song et al. (2015).

2.7 | SNPs and linkage analysis

One hundred eighty-one RILs were kept, and two were removed, due to a high percentage of heterozygosity and missing data, and 474 SNP markers were obtained after filtration for distortion, missing data and redundancy. Markers with greater than 20% missing data were removed for linkage map construction. A genetic linkage map was developed for Morden003/SMARC1N-PN1 using MapDisto v. 1.8.1 (Lorieux, 2012) with an r_{max} of 0.24, LOD_{min} of 3, and the Kosambi function. Marker order was optimized by the “order” and “ripple” functions. Linkage groups were assigned to a physical common bean chromosome (v2.1) by using BLASTn results from Phytozome 12 (https://phytozome.jgi.doe.gov/pz/portal.html#info?alias=Org_Pvulgaris). Linkage groups were illustrated using MapChart v. 2.30 (Voorrips, 2002).
2.8 | QTL mapping

QTL analyses were conducted using R/qtl v. 1.39–5 (Broman & Sen, 2009) in R v. 3.3.0 (R core Team 2016). Single-QTL genome scans were performed using multiple imputation with a scan interval of 1 cM (imputations = 1000; error probability = 0.001). Two-dimensional genome scans were performed using Haley-Knott regression with the thresholds based on the results of 1000 permutations at a 5% significance level to enable assessment of evidence for multi-QTL models involving additive or interacting loci.

2.9 | Allelic discrimination using PCR markers

Kompetitive Allele Specific PCR (KASP) assay of parent-specific SNP markers was performed with a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, Mississauga, ON) using a 10-μl reaction volume containing 4 μl of KASP genotyping mix (LGC Biosearch Technologies, Middlesex, UK), 1 μl of each primer (1.28 μM), which equates to 0.32 μM final concentration of each primer, and 5 μl of DNA (25 ng μl−1). The PCR conditions for KASP assay were 94°C for 15 min, followed by 10 cycles of touchdown PCR from 61°C to 55°C with 0.6°C decrease per cycle, and 25 cycles of 94°C for 20 s and 55°C for 60 s and a final read stage at 37°C for 1 min. Allele discrimination function of CFX Manager was used to carry out data analysis. Genomic sequence based SNP markers were designed using the genomes of G19833 (Schmutz et al., 2014) and BAT93 (Vasova et al., 2016) as references, as well as Great Northern US 1140 (NCBI Short Read Archive accession number: SRS2364194), Sanilac and SMARC1N-PN1 (Pandurangan et al., 2016), and Phaseolus coccineus accession G35346 (NCBI Short Read Archive accession number SRX1182014). Validation of these markers was performed by melting temperature (Tm)-shift SNP genotyping method (Wang et al., 2005).

2.10 | Statistical analysis

ANOVA was performed using SAS version 3.5 (SAS Institute Inc., Toronto, ON). Homogeneity of the variances of the sulfur amino acid data was evaluated by Bartlett’s test. Means were compared using Tukey’s range test. For agronomic data, ANOVA was performed with SPSS 11.0 (https://www.ibm.com/analytics/spss-statistics-software). Means were compared with Dunnett’s test.

3 | RESULTS

3.1 | Analysis of RILs

The SDS-PAGE of soluble protein extracts from mature seeds classified the RILs into four different groups according to their protein profile: 61 with the Morden003 profile for either phaseolin or lectins (MM), 59 with the Morden003 profile for phaseolin and with the SMARC1N-PN1 profile for lectins (MS), 38 with the SMARC1N-PN1 profile for phaseolin and with the Morden003 profile for lectins (SM) and 25 with SMARC1N-PN1 profile for both phaseolin and lectins (SS) (Dataset S1) (Viscarra-Torrico et al., 2021). Table S1 provides the high-density SNP genotyping information for the two parents and RILs, along with their genotype at the phaseolin and lectin loci, and the concentration of cysteine, methionine, cysteine + methionine, and S-methylcysteine. SNP genotyping information is also provided for parents of SMARC1N-PN1: the navy bean cultivars Sanilac and Great Northern US 1140, which carries an APA locus haplotype conferring erythroagglutinating phytohemagglutinin (pha-E) deficiency (Osborn & Bliss, 1985), along with three phaseolin-deficient P. coccineus genotypes (Pandurangan et al., 2016), the wild accession G12882 that contains the insecticidal lectin arcelin-1, and the related germplasm lines, SARC1 and SMARC1-PN1 (Osborn et al., 2003). For a subgroup of 26 lines, sulfur amino acids were determined separately at the SPARC BioCentre (Table S2). These data were excluded from the QTL analysis. However, they revealed similar contrasts between protein phenotypic groups (Viscarra Torrico, 2017). Amino acid data were missing for an additional subgroup of 14 lines.

3.2 | QTL analysis of sulfur amino acid traits

QTL mapping was performed to identify genomic loci contributing to the variation in sulfur amino acid concentration. A genetic map was built using the high-density genotyping information obtained with the BARCBean6K_3 array (Figure S1). The phenotypic markers, Lectin and Phaseolin, were mapped on Pv04 (~46 Mb) and Pv07 (~3 to 6 Mb), respectively (Figure 1a). Table 1 summarizes the information on the QTL identified in this study. Figure 1b shows the graph of logarithm of the odds (LOD) score versus position on the chromosome for cysteine, methionine and cysteine + methionine concentration. For cysteine, a single QTL was identified mapping to the phaseolin locus (R² = 38.7%). Figure 1c shows the contrast in cysteine concentration for genotypes differing at the marker representing the peak value of this QTL (ss715646455). For methionine, two QTL were identified corresponding to the phaseolin (R² = 21.1%) and lectin loci (R² = 29.2%). Genotypes homozygous for the SMARC1N-PN1 allele at the phaseolin locus and having the Morden003 allele at the lectin locus had increased methionine concentration (Figure 1d). This was enhanced in genotypes having SMARC1N-PN1 alleles at both loci. Similar results were observed for the sum of cysteine and methionine concentration (Figure 1e). No QTL was detected for S-methylcysteine concentration.

3.3 | Development of molecular markers for phaseolin deficiency

Analysis of high-density genotyping information of parental genotypes of SMARC1N-PN1 identified three phaseolin-linked markers having a unique allele in phaseolin-deficient P. coccineus accessions
These markers span the phaseolin locus at approximately 1 Mb of distance on either end. To identify SNP markers more closely linked with the trait, the intervals defined by the QTL were examined by comparing the genomic sequences of several genotypes investigated here as well as some reference genotypes. Two markers associated with the lectin locus on chromosome 4, and three markers associated with the phaseolin locus on chromosome 7 were identified. Figure 2 shows the position of these markers within the lectin and phaseolin loci. To validate markers, an analysis was performed to correlate the inheritance of the allele with the protein genotype identified by SDS-PAGE, for a selected group of RILs and their parents (3 MM, 3 MS, 3 SM and 3 SS). Table S3 summarizes the genotyping results and allele information for the markers. Table S4 lists the primers or sequence interval targeted for genotyping. Supplementary data document the detection of specific alleles at each locus (Dataset S2 to Dataset S5).

3.4 | Protein quality evaluation of selected RILs

A group of seven SS lines was selected for further analysis. Dataset S6 provides the average sulfur amino acid concentrations for each year and location, and the protein profile for individual plots. Based on these data, it was apparent that out of the seven SS lines, only two, 2-37 (from Pop2) and 3-84 (from Pop3), had a stable protein profile similar to SMARC1N-PN1 that remained consistent over years.
| Trait              | Chromosome | DF | SS      | LOD  | $R^2$ (%) | Probability (F) | Additive effect |
|--------------------|------------|----|---------|------|----------|-----------------|----------------|
| Cysteine           | 7          | 1  | 991.50  | 14.99| 38.70    | 2.22E-16        | 2.745          |
| Methionine         | 4          | 2  | 877.82  | 9.3  | 21.12    | 9.24E-10        | 2.164          |
|                    | 7          | 2  | 1215.16 | 12.24| 29.23    | 1.29E-12        | 2.650          |
| Cysteine + methionine | 4        | 2  | 1215.16 | 12.24| 29.23    | 1.29E-12        | 2.650          |
| Methionine         | 7          | 2  | 1215.16 | 12.24| 29.23    | 1.29E-12        | 2.650          |
| Cysteine + methionine | 4        | 2  | 1215.16 | 12.24| 29.23    | 1.29E-12        | 2.650          |

Note: DF: degrees of freedom; SS: sum of squares; LOD: logarithm of the odds; v.2 refers to the position in the version 2 of the reference genome, G19833 (Schmutz et al., 2014); NA: not available.

Abbreviation: QTL, quantitative trait loci.
and locations. These two lines were selected for protein quality evaluation.

Tables 2 and 3 summarize the agronomic characteristics of lines 2-37 and 3-84. Because each population represents a separate trial with parental controls, these lines are reported separately. Both lines were earlier maturing than SMARC1N-PN1, which is a significant improvement for adaptation. The seed yield and seed size for line 2-37 were similar to Morden003, whereas the seed yield and seed size were significantly smaller for line 3-84 compared to Morden003. Conversely, 2-37 exhibited a less desired floppy growth habit similar to SMARC1N-PN1; whereas 3-84 more closely resembled the compact determinate bush growth habit of Morden003. Data on agronomic properties of the remaining five lines in this sub-group of seven SS lines are provided in Viscarra Torrico (2017).

Table 4 reports the levels of sulfur amino acids averaged over both years and locations for lines 2-37, 3-84 and both parental lines. Cysteine levels in lines 2-37 and 3-84 were similar to those in SMARC1N-PN1, and significantly higher than in Morden003, by 32-37%. This was accompanied by a corresponding decrease in S-methylcysteine, by up to 38%, which was statistically significant in

![FIGURE 2](image)

**FIGURE 2** Position of SNP markers tracking storage protein deficiency on chromosome 4—APA locus (a) and chromosome 7—phaein locus (b). The figure depicts loci in the reference genotype G19833. Positions on the chromosome are indicated in base pairs.

**TABLE 2** Agronomic characteristics of line 2-37

| Genotype       | Agronomic traits |
|----------------|------------------|
|                | DTF (days) | DTM (days) | CHT (cm) | YLD (t ha⁻¹) | PDH (%) | HSW (g per 100 seeds) |
| Morden003      | 46 ± 2      | 95 ± 6     | 43 ± 5   | 1.9 ± 0.6    | 73 ± 16 | 17.2 ± 1.5             |
| SMARC1N-PN1    | 48 ± 2      | 109 ± 9    | 50 ± 10  | 1.9 ± 0.8    | 69 ± 15 | 17.2 ± 1.1             |
| 2-37           | 46 ± 3      | 101 ± 7    | 48 ± 8   | 1.9 ± 0.6    | 70 ± 10 | 18.3 ± 1.0             |

*Note: Genotypes were grown in the field 2014–2015 at London, ON and Morden, MB. Values are the average ± standard deviation. n = 3. DTF: days to flowering; DTM: days to maturity; CHT: canopy height; YLD = yield; PDH: pod height (percentage of pods above 5 cm from the ground); HSW: hundred seed weight. Means were not significantly different from those of Morden003 as determined by Dunnett’s test at p ≤ 0.05.*
TABLE 3 Agronomic characteristics of line 3-84

| Genotype     | DTF (days) | DTM (days) | CHT (cm) | YLD (t ha⁻¹) | PDH (%) | HSW (g per 100 seeds) |
|--------------|------------|------------|----------|--------------|---------|-----------------------|
| Morden003    | 46 ± 3     | 99 ± 6     | 41 ± 6   | 2.2 ± 0.9    | 76 ± 13 | 18.7 ± 1.2            |
| SMARC1-PN1   | 47 ± 2     | 109 ± 9    | 50 ± 9   | 2.2 ± 0.9    | 70 ± 14*| 17.9 ± 1.2            |
| 3-84         | 46 ± 3     | 97 ± 4     | 41 ± 6   | 1.6 ± 0.7*   | 67 ± 14*| 15.5 ± 0.7*           |

Note: Genotypes were grown in the field in 2014–2015 at London, ON and Morden, MB. Average ± standard deviation. n = 3.

*Significantly different from Morden003 at p ≤ 0.05 according to Dunnett’s test.

TABLE 4 Total sulfur amino acid levels in dry bean genotypes grown in London, ON and Morden, MB in 2014 and 2015

| Genotype     | S-Methylcysteine | Methionine | Cysteine |
|--------------|-------------------|------------|----------|
| Morden003    | 14.6ᵃ             | 11.7       | 16.5ᵃ    |
| SMARC1-PN1   | 10.2ᵇ           | 12.8       | 23.0ᵇ    |
| 2-37         | 9.92ᵇ           | 13.8       | 21.7ᵇ    |
| 3-84         | 9.38ᵇ           | 13.1       | 22.6ᵇ    |
| CV (%)       | 18.8             | 17.0       | 6.98     |
| F-value      | 5.32*            | 0.64       | 17.0**   |

Note: Data are expressed in nmol per mg seed weight. CV: coefficient of variance; values with the same letter are not significantly different according to Tukey’s range test at p ≤ 0.05 for S-methylcysteine and p ≤ 0.001 for cysteine; Tukey’s range test was performed after logarithmic transformation of the data for S-methylcysteine and cysteine; *ANOVA p value ≤ 0.05. **ANOVA p value ≤ 0.0001.

Administration for protein content claims on foods. As expected, sulfur amino acids were the most limiting. However, there was a relatively large increase in line 2-37 as compared with Morden003, for both amino acid score of sulfur amino acids, by 43% and PDCAAS, by 37%. There was a notable increase in amino acid score for tryptophan, by 52%, whereas this value was decreased for leucine by 12%. Line 3-84 had smaller increases in sulfur amino acid score (ca. 10%) or PDCAAS (6%) as compared with Morden003.

4 | DISCUSSION

The results of QTL analyses performed in this study clarify the effects of the phaseolin and lectin loci on variation in cysteine and methionine concentrations in the Morden003 × SMARC1-PN1 RIL population. QTL were identified, even though amino acid data for some of the lines were either missing or excluded due to the fact they were obtained on a different analytical platform, and therefore were not directly comparable with those of the majority of the RILs. The phaseolin locus was the only QTL associated with cysteine concentration (Table 1 and Figure 1b). Phaseolin-deficiency led to increased methionine concentration, with an additive effect of lectin deficiency (Figure 1d).

The SNP markers identified in this study may be useful to test genetic material for phaseolin or lectin deficiency, and to follow the introgression of the traits (Figure 2). Parent-of-origin specific SNPs present on the genotyping array are able to track the segregation of phaseolin deficiency, although they are relatively distant, approximately 1 Mb on either side of the corresponding locus (Table S3). Fine mapping the QTL regions by comparing genomic sequences among relevant genotypes enabled the development of improved markers, located within or closer to the loci (Figure 2). The lectin markers are located upstream, about 10 kb away from the PDLEC2 gene. The phaseolin markers are located downstream from the r-phaseolin genes, which are presumed to be absent in SMARC1-PN1. Marker S07_514621 falls between two copies of the β-phaseolin gene. The other two markers are further away, about 50-kb downstream.

The major objective of this study was to generate germplasm lines having increased sulfur amino acid levels, with adaptation to short season conditions in Manitoba. Multiple challenges were encountered. The first one was associated with the stability of the protein profile. Out of seven SS lines, only two had a particularly
stable protein profile over 2 years and locations (Dataset S6). Factors that may have affected the stability of protein composition include outcrossing, although its frequency is low in common bean (Ibarra-Perez et al., 1997). To face this potential problem, physical isolation might be required to grow storage protein deficient common bean genotypes, like what is practiced with waxy corn, for example. Environmental variability is another possible factor that may have influenced the stability of the protein profile. It is worth noting that SMARC1N-PN1 is not completely devoid of phaseolin, due to the presence of a residual, functional copy of a β-phaseolin gene. A 20-fold decrease in β-phaseolin accumulation was associated with a polymorphism converting a proximal G-box to an ACGT motif in the promoter of β-phaseolin (Pandurangan et al., 2016). The presence of this residual phaseolin gene may contribute to variation in protein profile in the RILs. Tracking phaseolin and lectin deficiency at different stages of breeding, using the PCR markers identified in this study (Figure 2), is likely to facilitate the recovery of pure genotypes, having a stable protein profile.

The results of amino acid analyses of genotypes grown over 2 years at two locations confirmed prior results linking storage protein deficiency with increased cysteine concentration (Table 4) (Taylor et al., 2008). However, the increase in cysteine concentration (approximately 30%–40%) was not as high as previously reported in the initial study (up to 70%, when compared to SARC1 as the reference genotype). To maximize the concentration of sulfur amino acids, it may be necessary to apply sulfate fertilizer. This was only done at the Morden site, a common practice in Manitoba, but not generalized in Ontario. Cysteine concentration was previously shown to positively respond to sulfate nutrition in SMARC1N-PN1, under controlled conditions (Pandurangan, Sandercock, et al., 2015). In the present study, line 2-37 displayed a substantial increase in PDCAAS, based on in vitro digestibility assay (Table 6). The increased protein quality for this line is attributable to higher sulfur amino acid levels, while having a slightly lower protein concentration. The differences in crude protein concentration, as determined by elemental analysis, were consistent with those determined for the sum of total amino acids, measured after acid hydrolysis. The increase of in vitro PDCAAS measured for line 3-84 as compared with Morden003 was relatively marginal, equal to only 6.5%. This reveals how potentially difficult it may be to implement a strategy based on storage protein deficiency to improve protein quality. Tryptophan is the second most limiting amino acid in pulses after sulfur amino acids (Nosworthy et al., 2017). The amino acid score for tryptophan was substantially increased, particularly in line 2-37 as compared with Morden003. This observation is of interest for protein quality improvement. Among other nutritionally essential amino acids, leucine concentration was decreased in both lines, while lysine concentration was decreased in line 3-84.

Common bean genotypes lacking phaseolin and major lectins were initially developed as part of an effort to introduce and favor the accumulation of arcelins, a group of insecticidal lectins from wild common bean, active against several bean weevils, which are pests of stored grain (Osborn et al., 2003). In general, genotypes lacking...
phaseolin and major lectins have similar agronomic characteristics as regular genotypes (Hartweck & Osborn, 1997; Osborn & Bliss, 1985). This is consistent with the findings from this study. The lectin deficiency present in SMARC1N-PN1 does not involve a complete lack of lectins. Rather, the absence of erythroagglutinating phytohemagglutinin encoded by pha-E, leucoagglutinating phytohemagglutinin encoded by pha-L, and lectin encoded by lec4-B17 is compensated by high levels of leucoagglutinating phytohemagglutinin encoded by PDLEC2, α-amylase inhibitor 1, α-amylase inhibitor-like protein and mannos lectin FRIL (Pandurangan et al., 2016). Kusolwa et al. (2016) recently introduced an arcelin from tepary bean (Phaseolus acutifolius) into a phaseolin and lectin-deficient background. They found that the new weevil-resistant lines had higher levels of threonine (34%–44%), cysteine (15%–20%), methionine (0%–11%), lysine, and valine, whereas the levels of isoleucine and leucine stayed the same or were decreased, similar with the present study. Giuberti et al. (2019) recently examined correlations between variations in protein composition and nutritional characteristics of common bean. They reported a strong positive correlation between genotypes lacking phaseolin and having high levels of α-amylase inhibitor, as in the present study, and the accumulation of iron and zinc. This could be of further interest from a biofortification point of view.

In conclusion, a germplasm line, designated HS 2-37, has been generated, that exhibits a substantial increase in protein quality as compared with its commercial parent cultivar. This property might be interesting from a nutritional standpoint, or for specific food applications, since it is also relevant to nutritional claims on protein content. In addition to testing agronomic characteristics, breeding for this trait likely requires stringent controls to ensure genetic purity, using molecular markers, as well as monitoring for seed quality characteristics, including protein concentration and essential amino acid profiles.

| TABLE 6 Protein digestibility corrected amino acid scores of common bean lines (1991 reference pattern for all age groups except infants) |
|-----------------------------------------------|
| **Reference pattern** (mg per g crude protein) | **Histidine** (His) | **Isoleucine** (Ile) | **Leucine** (Leu) | **Lysine** (Lys) | **M + C** | **P + T** | **Threonine** (Thr) | **Tryptophan** (Trp) | **Valine** (Val) |
| Morden003                                      | 1.28               | 1.44               | 1.13               | 0.99               | 0.623     | 1.33       | 1.13               | 0.95               | 1.26             | 0.623     | 72.6 | 0.452 |
| 2-37                                           | 1.36               | 1.38               | 1.01               | 1.01               | 0.891     | 1.29       | 1.38               | 1.45               | 1.35             | 0.891     | 69.5 | 0.619 |
| 3-84                                           | 1.17               | 1.33               | 1.04               | 0.91               | 0.684     | 1.25       | 1.28               | 1.16               | 1.25             | 0.684     | 70.4 | 0.482 |

*Note: PD, Protein Digestibility; PDCAAS, Protein Digestibility-Corrected Amino Acid Score. PDCAAS is calculated as the product of AA score and percent PD (FAO, 2013). *

* aM + C = methionine + cysteine.

* bP + T = phenylalanine + tyrosine.

ACKNOWLEDGMENTS

We are grateful to Molly Welsh and staff at the Germplasm Resources Information Network of the United States Department of Agriculture-Agricultural Research Service, Western Regional Plant Introduction Station, Pullman, WA, for providing seeds of G12882 and Great Northern 1140. We thank Kathleen Hill for acting as RCV-T’s supervisor at the Department of Biology of the University of Western Ontario, Alex Molinar for help with figures, and Vanessa Ma, for assistance with DNA extraction. We are grateful to Michael Nickerson from the University of Saskatchewan for providing the protocol to test for in vitro protein digestibility. We thank Brett Hill, from the Lethbridge Research and Development Centre of Agriculture and Agri-Food Canada, for elemental analysis. Funding information: Manitoba Pulse and Soybean Growers and Pulse Science Cluster 2 project of Agriculture and Agri-Food Canada (Agriculture and Agri-Food Canada project IDs J-000583 and J-000553).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

RCV-T, RLC, PNM, AH and FM contributed to conceptualization; RCV-T, AP, ASG, BZ and SP to investigation; ASG and MD to formal analysis; QS, PBC, PNM, and JDH contributed resources; RCV-T, AP, ASG, BZ, AH, and FM contributed to writing—original draft; SP, QS, JDH, PNM, AF, and FM contributed to writing—review and editing; FM and AH participated to supervision; AH was responsible for funding acquisition.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available through the Scholars Portal Dataverse at https://doi.org/10.5683/SP2/JL7Y2I. Lines developed through this study have been deposited at the Plant Gene Resources of Canada under the following names and accession numbers: HS 2-37 (CN 120262) and HS 3-84 (CN 120263).

ORCID

Frédéric Marsolais https://orcid.org/0000-0003-3914-141X

REFERENCES

Broman, K. W., & Sen, S. (2009). A guide to QTL mapping with R/qtl. New York, NY: Springer Science+Business Media. https://doi.org/10.1007/978-0-387-92125-9
De Ron, A. M., Papa, R., Bitocchi, E., González, A. M., Debouch, D. G., Brick, M. A., Fourie, D., Marsolais, F., Beaver, J., Geffroy, V., McClean, P., Santalla, M., Lozano, R., Yuste-Lisbona, F. J., & Casquerio, P. A. (2015). Common bean. In A. M. De Ron (Ed.), *Grain legumes* (pp. 1–36). New York: Springer Science+Business Media. https://doi.org/10.1007/978-1-4939-2797-5_1

Diapari, M., Pandurangan, S., Song, Q., Cregan, P., Miklas, P. N., & Marsolais, F. (2016). High-density SNP genotyping of lines with different storage protein composition reveals sites of interspecific introgression and markers associated with phaseolin or lectin deficiency. *Annual Report of the Bean Improvement Cooperative*, 59, 1–2.

FAO. (2013). Dietary protein quality evaluation in human nutrition. Report of an FAO Expert Consultation, 31 March-2 April 2011, Auckland, New Zealand. FAO Food and Nutrition Paper. 92. Rome.

FAO/WHO. (1991). Protein quality evaluation. Report of the Joint FAO/WHO Expert Consultation. FAO Food and Nutrition Paper. 51. Rome.

Freyre, R., Skroch, P. W., Geffroy, V., Adam-Blondon, A. F., Shirmohammadi, A., Johnson, W. C., Ilcà, V., Nodari, R. O., Pereira, P. A., Tsai, S. M., Tohme, J., Dron, M., Nienhuis, J., Vallejos, C. E., & Gepts, P. (1998). Towards an integrated linkage map of common bean. 4. Development of a core linkage map and alignment of RFLP maps. *Theoretical and Applied Genetics*, 97(5–6), 847–856. https://doi.org/10.1007/s001220050964

Gibert, G., Tava, A., Mennera, G., Pecetti, L., Masoero, F., Sparvcoli, F., ... Campion, B. (2019). Nutrients and antinutrients' seed content in common bean (*Phaseolus vulgaris* L.) lines carrying mutations affecting seed composition. *Agronomy*, 9(6), 317. https://doi.org/10.3390/agronomy9060317

Hartweck, L. M., & Osborn, T. C. (1997). Altering protein composition by genetically removing phaseolin from common bean seeds containing arcelin or phytohemagglutinin. *Theoretical and Applied Genetics*, 95(5–6), 1012–1017. https://doi.org/10.1007/s001220050655

Ibarra-Perez, F. J., Ehdai, B., & Waines, J. G. (1997). Estimation of outcrossing rate in common bean, *Phaseolus* Cv. Crop Science, 37(1), 60–65. https://doi.org/10.2135/cropsci1997.0011183X003700010009x

Technical Committee ISO/TC 34/SC 10. (2016). Animal feeding stuffs—Determination of tryptophan content ISO 13904. Geneva: International Organization for Standardization.

Jafari, M., Rajabzadeh, A. R., Tabtabaei, S., Marsolais, F., & Kemp, G. (2004). Morden003 navy bean. *Canadian Journal of Plant Science*, 84(1), 231–233. https://doi.org/10.4141/P03-059

Joshi, J., Pandurangan, S., Diapari, M., & Marsolais, F. (2017). Comparison of gene families: seed storage and other seed proteins. In M. Pérez de la Vega, M. Santalla, & F. Marsolais (Eds.), *Phaseolus vulgaris*—protein fraction produced using a solvent-free method. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-63552-6_10

Kusolwa, P. M., Myers, J. R., Porch, T. G., Trukhina, Y., González-Velázquez, A., & Beaver, J. S. (2016). Registration of AO-1012-29-3-3A red kidney bean germplasm line with bean weevil, *BCMV*, and *BCMNV* resistance. *Journal of Plant Registrations*, 10(2), 149–153. https://doi.org/10.3198/jpr2015.10.0064cr

Liao, D., Pajak, A., Karcz, S. R., Chapman, B. P., Sharpe, A. G., Austin, R. S., Datla, R., Dhaubhadel, S., & Marsolais, F. (2012). Transcripts of Sulphur metabolic genes are co-ordinately regulated in developing seeds of common bean lacking phaseolin and major lectins. *Journal of Experimental Botany*, 63(17), 6283–6295. https://doi.org/10.1093/jxb/ers280

Lorieux, M. (2012). MapDisto: Fast and efficient computation of genetic linkage maps. *Molecular Breeding*, 30(2), 1231–1235. https://doi.org/10.1007/s11032-012-9706-y

Marsolais, F., Pajak, A., Yin, F., Taylor, M., Gabriel, M., Merino, D. M., Ma, V., Kameka, A., Vijayan, P., Pham, H., Huang, S., Rivoal, J., Betk, K., Hernandez-Sebastiá, C., Liu, Q., Bertrand, A., & Chapman, R. (2010). Proteomic analysis of common bean seed with storage protein deficiency reveals up-regulation of sulfur-rich proteins and starch and rafinose metabolic enzymes, and down-regulation of the secretory pathway. *Journal of Proteomics*, 73(8), 1587–1600. https://doi.org/10.1016/j.jprot.2010.03.013

Montoya, C. A., Lallès, J. P., Beebe, S., & Lettere, P. (2010). Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (*Phaseolus vulgaris*). *Food Research International*, 43(2), 443–449. https://doi.org/10.1016/j.foodres.2009.09.040

Mündel, H.-H., Kiehn, F. A., Huang, H. C., Conner, R. L., Saindon, G., & Kemp, G. (2004). *Morden003* navy bean. *Canadian Journal of Plant Science*, 84(1), 231–233. https://doi.org/10.4141/P03-059

Osborn, T. C., & Bliss, F. A. (1986). Bean arcelin 2. Genetic variation inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris*. *Theoretical and Applied Genetics*, 71(6), 847–855. https://doi.org/10.1007/BF00276428

Osborn, T. C., Hartweck, L. M., Harmsen, R. H. V., Vogelzang, R. D., Kmieciek, K. A., & Bliss, F. A. (2003). Registration of *Phaseolus vulgaris* genetic stocks with altered seed protein compositions. *Crop Science*, 43(4), 1570–1571. https://doi.org/10.2135/cropsci2003.1570

Pandurangan, S., Pajak, A., Rintoul, T., Beyerat, R., Hernandez-Sebastiá, C., Brown, D. C., & Marsolais, F. (2015). Soybean seeds overexpressing asparaginase exhibit reduced nitrogen concentration. *Physiologia Plantarum*, 155, 126–137. https://doi.org/10.1111/ppl2.12341

Pandurangan, S., Sandercock, M., Beyerat, R., Conn, K. L., Hou, A., & Marsolais, F. (2015). Differential response to sulfur nutrition of two common bean genotypes differing in storage protein composition. *Frontiers in Plant Science*, 6, 92. https://doi.org/10.3389/fpls.2015.00092

Pandurangan, S., Diapari, M., Yin, F., Munholland, S., Perry, G., Chapman, B. P., ... Marsolais, F. (2016). Genomic analysis of storage protein deficiency in genetically related lines of common bean (*Phaseolus vulgaris*). *Frontiers in Plant Science*, 7, 389. https://doi.org/10.3389/fpls.2016.00389

Pandurangan, S., & Marsolais, F. (2017). Analysis of phaseolin copy number by quantitative PCR. *Annual Report of the Bean Improvement Cooperative*, 60, 107–108.

Planta, J., & Messing, J. (2017). Quality protein maize based on reducing sulfur in leaf cells. *Genetics*, 207(4), 1687–1697. https://doi.org/10.1534/genetics.117.300288

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abemathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., ... Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of domestications. *Nature Genetics*, 46(7), 707–713. https://doi.org/10.1038/ng.3008

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., Osorno, J. M., Schmutz, J., Jackson, S. A., McClean, P. E., & Cregan, P. B. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and
genomic applications in common bean. G3, 5(11), 2285–2290. https://doi.org/10.1534/g3.115.020594

Talbot, D. R., Adang, M. J., Slightom, J. L., & Hall, T. C. (1984). Size and organization of a multigene family encoding phaseolin, the major seed storage protein of Phaseolus vulgaris L. Molecular and General Genetics, 198(1), 42–49. https://doi.org/10.1007/BF00328698

Taylor, M., Chapman, R., Beyaert, R., Hernández-Sebastià, C., & Marsolais, F. (2008). Seed storage protein deficiency improves sulfur amino acid content in common bean (Phaseolus vulgaris L.): Redirection of sulfur from γ-glutamyl-S-methyl-cysteine. Journal of Agricultural and Food Chemistry, 56(14), 5647–5654. https://doi.org/10.1021/jf800787y

Viscarra Torrico, R. C. (2017). The genetic improvement of protein quality in common bean (Phaseolus vulgaris L.). University of Western Ontario. Electronic thesis and dissertation repository. 4865. https://ir.lib.uwo.ca/etd/4865

Viscarra-Torrico, R. C., Pajak, A., Soler Garzón, A., Zhang, B., Pandurangan, S., Diapari, M., Song, Q., Conner, R. L., Miklas, P. N., Hou, A., & Marsolais, F. (2021). Phenotyping and genotyping of recombinant inbred lines and related parents of common bean (Phaseolus vulgaris L.) segregating for seed protein composition. Scholars Portal Dataverse. https://doi.org/10.5683/SP2/JL7Y2I

Vitale, A., & Bollini, R. (1995). Legume storage proteins. In J. Kigel & G. Galili (Eds.), Seed development and germination (pp. 73–102). New York: Marcel Dekker.

Vlasova, A., Capella-Gutierrez, S., Rendon-Anaya, M., Hernandez-Onate, M., Minoche, A. E., Erb, I., ... Guigo, R. (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. Genome Biology, 17(1), 32. https://doi.org/10.1186/s13059-016-0883-6

Voomips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. Journal of Heredity, 93(1), 77–78. https://doi.org/10.1093/fered/93.1.77

Wang, J., Chuang, K., Ahluwalia, M., Patel, S., Umblas, N., Mirel, D., Higuchi, R., & Germer, S. (2005). High-throughput SNP genotyping by single-tube PCR with tm-shift primers. BioTechniques, 39(6), 885–893. https://doi.org/10.2144/000112028

Wiggins, A. K. A., Anderson, G. H., & House, J. D. (2018). Research and regulatory gaps for the substantiation of protein content claims on foods. Applied Physiology, Nutrition, and Metabolism, 44(1), 95–98. https://doi.org/10.1139/apnm-2018-0429

Yin, F., Pajak, A., Chapman, R., Sharpe, A., Huang, S., & Marsolais, F. (2011). Analysis of common bean expressed sequence tags identifies sulfur metabolic pathways active in seed and sulfur-rich proteins highly expressed in the absence of phaseolin and major lectins. BMC Genomics, 12, 268. https://doi.org/10.1186/1471-2164-12-268

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Viscarra-Torrico, R. C., Pajak, A., Garzón, A. S., Zhang, B, Pandurangan, S., Diapari, M., Song, Q, Conner, R. L., House, J. D., Miklas, P. N., Hou, A., & Marsolais, F. (2021). Common bean (Phaseolus vulgaris L.) with increased cysteine and methionine concentration. Legume Science, e103. https://doi.org/10.1002/leg3.103