GERMPLASM EVALUATION OF COMMON BEANS FOR END-USE QUALITY: SEED COAT COLOR AND COOKING TIME OF YELLOW BEANS AND PASTE QUALITIES OF MICHIGAN BEAN VARIETIES

By

Rie Sadohara

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics, and Biotechnology—Crop and Soil Sciences—Doctor of Philosophy

2021

ABSTRACT

GERMPLASM EVALUATION OF COMMON BEANS FOR END-USE QUALITY: SEED COAT COLOR AND COOKING TIME OF YELLOW BEANS AND PASTE QUALITIES OF MICHIGAN BEAN VARIETIES

By

Rie Sadohara

Consumption of dry beans is low in the U.S. and the genetic improvement of end-use quality traits of dry beans and the development of new uses for beans offer potential opportunities to increase consumption. This research aimed to 1) characterize a yellow bean diversity panel for seed coat color and cooking time, 2) investigate the G×E and inheritance of seed coat color and cooking time of the diversity panel via genome-wide association (GWA), 3) evaluate the suitability of MI-bred bean varieties for sweet paste qualities, and 4) inform bean breeders of the food industry's needs for bean flour. A Yellow Bean Collection with 295 lines grown in MI and NE were evaluated for seed coat color, post-harvest darkening, and cooking time. Machine-learning was used to develop a procedure to automatically exclude hilum and corona areas in bean images and extract L*a*b* color values from seed coat. Color values, post-harvest darkening, hilum, and corona colors mapped to the ground factor *P* gene. Cooking time evaluation identified 20 lines that cooked within 20 min in MI and five lines that cooked within 31 min in higher-altitude NE. GWA identified a polygalacturonase gene as a candidate gene for cooking time. G×E was significant for seed coat color and cooking time, and MI-grown seeds were generally darker and faster-cooking than NE-grown. Six white bean varieties and one cranberry genotype were evaluated for sweet bean paste qualities: paste yield, color, stickiness, and flavor. Large-seeded white varieties performed similar or higher for many of the quality traits than Hime, an industry standard, showing their potential as a novel food application in the U.S. An electronic survey was conducted targeting U.S. food

industry professionals who use pulse and/or wheat flour. Seventy-five valid responses were collected, and flavor and functionality rather than lectins or gluten contamination were the most important challenges in utilizing bean flour. These studies will be useful in selecting germplasm for favorable qualities to consumers such as bright yellow color, non-darkening after storage, fast cooking time, and high bean paste and bean flour qualities.

ACKNOWLEDGMENTS

First and foremost, I would like to express sincere gratitude to my primary advisor, Dr. Karen Cichy for her encouragement, constructive feedback, and unwavering support throughout my doctoral research. Her scientific expertise, kindness, and receptiveness to new ideas makes her an ideal supervisor, and it has truly been a privilege to receive her guidance. I am grateful to the rest of my committee, Dr. James Kelly, Dr. Dechun Wang, and Dr. Donna Winham who dedicated their time to share their insights and guide me with their expertise. I am grateful to the former and current Cichy lab and Kelly lab members, other graduate students, NRT-IMPACTS trainers and trainees, and faculty and staff at the Department of Plant, Soil, and Microbial Sciences for their support, collaborations, and friendships. My heartfelt appreciation goes to my former supervisors and colleagues at Sanwa Shurui Co., Ltd. as well as Dr. Jason Eglinton and Ms. Sophia Degner at the University of Adelaide who all contributed to my personal and professional growth and helped me lay a foundation for pursuing an advanced degree. I would also like to thank my family – my parents, grandparents, sister, and uncles and aunts who supported me and believed in me throughout this journey. I have been very fortunate to pursue my Ph.D. degree at Michigan State University, and this work would not have been possible without all the people mentioned above.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

THE *PHASEOLUS VULGARIS* YELLOW BEAN COLLECTION: GENETIC DIVERSITY AND CHARACTERIZATION FOR COOKING TIME

[Submitted for publication in Genetic Resources and Crop Evolution]

The *Phaseolus vulgaris* Yellow Bean Collection: Genetic diversity and characterization for cooking time

Rie Sadohara[1], Paulo Izquierdo[1], Filipe Couto Alves[2], Timothy Porch[3], James Beaver[4], Carlos A. Urrea[5], Karen Cichy[1,6]

1. Department of Plant, Soil, and Microbial Sciences, Michigan State University, MI, USA
2. Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA
3. Tropical Agriculture Research Station, USDA-ARS, Mayaguez, PR, USA
4. Department of Agroenvironmental Sciences, University of Puerto Rico, Mayaguez, PR, USA
5. Panhandle Research & Extension Center, Scottsbluff, NE, USA
6. Sugarbeet and Bean Research Unit, USDA-ARS, East Lansing, MI, USA

Abstract

Common bean (*Phaseolus vulgaris* L.) is a nutrient-rich food, but its long cooking times hinder its wider utilization. The Yellow Bean Collection (YBC) was assembled with 295 genotypes from sources globally to assess the genetic and phenotypic diversity for end-use quality traits in yellow-colored beans. The panel was genotyped via Genotyping-By-Sequencing (GBS) identifying over 2,000 SNPs. Through population structure analyses with the GBS markers, the YBC was determined to be 69% Andean, 26% Middle American, and 5% admixture. The YBC was grown in two major bean production regions in the US, Michigan and Nebraska over two years. The genotypes exhibited wide diversity in days to flower, seed weight, water uptake, and cooking time. Notably, the cooking times of the YBC ranged from 15-210 min among environments and years. The cooking time varied more widely in NE with many more genotypes exhibiting hardshell than

in MI. Fast-cooking genotypes were identified with various yellow colors; 20 genotypes cooked within 20 min in MI, and five genotypes cooked within 31 min at a higher altitude in NE. Water uptake and cooking time were significantly affected by the environment, resulting in a genotype-by-environment interaction. Significant SNPs for cooking time were identified with genome-wide association analyses and a polygalacturonase gene on chromosome Pv04 was considered to be a candidate gene for hardshell and cooking time. The genotypic and phenotypic variability, fast-cooking genotypes, and the associated SNPs of the YBC will lay the foundation for utilizing yellow beans for breeding and genetic analyses.

Introduction

Common bean (*Phaseolus vulgaris* L.) is rich in diversity for seed sizes, shapes, colors, and patterns. One interesting color class within *P. vulgaris*, is yellow. Yellow beans occur in many shades and sizes (**Figure 1.1**), and they make up important market classes in Latin America, the Caribbean, and Africa (Kilimo Trust, 2012; Voysest, 2012; Wortmann et al., 1998). The premier source of yellow colored beans is the Peruvian coast, where they have been grown since ancient times (Voysest, 2012). Including some unique classes only found in Peru, there are at least a dozen market classes of yellow beans produced in Latin America. Yellow beans are also important in African countries including Angola, Tanzania, and Kenya (Buruchara et al., 2011). In fact, yellow beans fetch higher prices than other seed types in Zambia (Sichilima et al., 2016), and they are the most preferred type in Uganda (Kilimo Trust, 2012). A few of the most important yellow market classes include Manteca, Canario, and Mayocoba in Latin America, and Njano in Tanzania (Sones, 2015).

Manteca is a market class with pale lemon-yellow seed coat and medium-sized seeds. Mantecas have been preferred by consumers in Chile because they were easier to digest and were more expensive due to this added benefit (M. J. Bassett, 1999; Leakey, 1992). The consumers' perception of low flatulence may be explained by the observations that Manteca beans are free of tannins, which can decrease starch and protein digestibility (Beninger et al., 1998), and low in indigestible proteins, starch, and insoluble dietary fiber (Engleright et al., 1999; Hooper et al., 2016; Hosfield et al., 1998). Beans of lemon yellow color have other advantages such as good taste and texture (Beninger et al., 1998; Hosfield et al., 1998). Moreover, some Manteca beans have been found to be fast-cooking and high in bioavailable iron (Cichy, Wiesinger, et al., 2015; J. Wiesinger et al., 2018) due to their unique polyphenolic profiles with high levels of iron absorption promoters and low levels of inhibitors (Hart et al., 2020). Therefore, Manteca beans have offered multiple end-use and nutritional attributes that benefit consumers.

Peruvian Canario has a bright yellow seed coat color similar to a canary bird with no hilum ring color and has a long history of cultivation in Peru (Voysest, 2012). It was introduced to Mexico in the 1970s and was given the name Peruano in Mexico to distinguish it from the Canario type that existed in Mexico, which has a duller yellow seed coat and a dark hilum ring. Peruano was crossed with Mexican Canario to produce varieties that have the same seed coat color as the bright yellow Peruano. One of the resultant varieties, 'Azufrado Pimono 78', is also called Mayocoba, and Mayocoba became the name for the Peruano seed type in the U.S. (M. J. Bassett, 2002; Voysest, 2012). Mayocoba beans have been involved in a highly controversial U.S. patent granted to a Peruano type yellow bean called Enola in 1999, which claimed exclusive monopoly rights for the bright yellow seed color (Proctor, 1999). DNA fingerprinting concluded that Enola was most likely a selection of yellow beans brought back from Mexico, involving no 'development'

of the color, and genetically identical to Azufrado Peruano 87 (Pallottini et al., 2004). The patent was cancelled after a protracted legal battle, but this incident discouraged U.S. bean breeders to invest as much effort in the improvement of the Mayocoba bean class as compared to other market classes. However, Mayocoba beans have been included in varietal trials for many years, and some Mayocoba-type varieties were recently released including Yellowstone and Patron (US Department of Agriculture, Agricultural Marketing Service 2019; Kelly et al., 2021).

Yellow beans are important market classes in Eastern and Southern Africa, where beans are produced and consumed as a dietary staple (Broughton et al., 2003; Wortmann et al., 1998). Yellow beans in Africa vary in shape, size, and color with regional preferences by consumers (Sichilima et al., 2016; Tumeo et al., 2017; Wortmann et al., 1998). Green-yellow shaded beans are referred to as Njano Kijani and Kijivu in Tanzania (Dr. Susan Nchimbi-Msolla, Sokoine University of Agriculture, Tanzania, personal communication), and Njano is a preferred type for home consumption in Tanzania (Sones, 2015). The yellow bean class in Angola, Manteiga, is a popular seed class composed of several seed coat colors, ranging from cream to yellow, often grown or marketed in mixtures. Yellow, tan, and brown seed types are particularly popular in Angola and account for 11% of seed produced in the region including Eastern and Southern Africa (Wortmann et al., 1998). However, information is limited about preferred seed types and the history of yellow beans in other African countries. A thorough and updated review is needed for the diversity of yellow beans in Africa similar to Voysest (2012) which reviewed all the yellow beans present in Latin America.

One of the quality characteristics associated with some market classes of yellow beans is fast cooking times (Mishili et al., 2011). Cooking time is the length of time beans take to be cooked and palatable. Beans typically take 30 min - 2 hours to be cooked, but beans with shorter cooking

times are demanded by consumers and could lead to increased consumption in countries where beans are eaten less frequently (Aseete et al., 2018; Brouwer et al., 1996; Smith et al., 2013; J. Wiesinger et al., 2018). In order to improve any trait including cooking time through plant breeding, wide genetic diversity and a high heritability for the trait are desirable.

A wide genetic variability in cooking time has been observed in common beans, and yellow beans were one of the fastest cooking types identified in an Andean bean diversity panel (ADP) (Cichy, Wiesinger, et al., 2015). In addition, yellow beans are found in both of the two genepools of common bean, Andean and Middle American (MA) (Voysest, 2012), making inter- or intra-genepool crosses an option for plant breeders. The heritability of cooking time is estimated to be moderate to high (Cichy et al., 2019; Elia et al., 1997; Garcia et al., 2012; Jacinto-Hernandez et al., 2003; Katuuramu et al., 2020). The high genetic variability and heritability highlights the possibility of improving (shortening) cooking times of beans, and yellow beans will be a useful material to study and exploit for the fast-cooking trait. In addition to genetic architecture, cooking time of beans can also be impacted by the production environment (A. Bassett, Kamfwa, et al., 2021; Berry et al., 2020; Cichy et al., 2019). Therefore, it is necessary to assess the environmental effect and genotype x environment interaction on the cooking time performance of bean germplasm.

The objectives of this research were to 1) describe the genetic diversity of a Yellow Bean Collection (YBC) of 295 yellow beans (*P. vulgaris*) assembled with germplasm from the Americas, Europe, and Africa; 2) evaluate bean water uptake and cooking time when the YBC was grown in two major bean producing regions in the U.S.; and 3) investigate the genetic control of water uptake and cooking time and G x E effects.

Materials and methods

Plant material

A collection of 295 *Phaseolus vulgaris* L. accessions with various yellow seed colors were assembled from global sources (**Table 1.1**). This diversity panel is called the Yellow Bean Collection (YBC). A set of 14 non-yellow genotypes and commercial varieties (white kidney, navy, cranberry, red mottled, light red kidney, and dark red kidney) were evaluated as controls for this panel: some with good agronomic adaptation to Michigan growing conditions (navy, dark red kidney, and light red kidney), and some previously characterized for cooking time G x E (Cichy et al., 2019). The characteristics of the control genotypes are shown in **Table S1.6**.

Field design

The YBC lines and 14 control varieties were grown at Michigan State University Montcalm Research Station (43°21.2'N, 85°10.6'W, 284 m elevation) in 2018 and 2019, and at the Scottsbluff (41°53.6’N, 103°40.7’W, 1200 m elevation) and Mitchell, Nebraska, Ag Labs (41°56.6’N, 103°41.9’W, 1240 m elevation), in 2018 and 2019, respectively. The local standard agricultural practices were followed in terms of fertilizer, herbicide, and pesticide applications as described by Kelly et al. (2018) and Kelly et al. (2019) in all the environments. Sets of 210-265 genotypes were planted and phenotyped for days to flower, 100-seed weight, water uptake, and cooking time (**Table S1.7**). Fewer genotypes were planted in Nebraska due to the initial shortage of seed of the 27 lines from Haiti. Each genotype was planted in a randomized complete block design with two field replications. At the Montcalm Research Station in Michigan, each plot consisted of four 6.1-m rows with 0.51-m row width. The central and the outer rows of each plot were sown with seeds of the YBC genotypes and a dark red kidney bean, respectively, with sowing density of around 13

seeds per linear meter. Forty-five lines were sown at a reduced rate (25 seeds/row) in 2018 due to a shortage of seed. Rainfall was supplemented with an overhead irrigation applied 10 times throughout the growing season with 13.7 mm (5.4 inches) in total in 2018 and 12 times with 18.4 mm (7.25 inches) in total in 2019. At maturity, all bean plants in the center 4.6-m section of the center two rows were hand-pulled, and the seeds were harvested using a Hege plot thresher. Harvested seeds were dried at room temperature, cleaned to remove chaff, stones, and plant debris by using a seed cleaner (Clipper Office Tester, AT Ferrell Company Inc., Bluffton IN, USA), hand-cleaned to remove gravel and damaged or foreign seeds, and stored at a low-temperature and low-humidity storage chamber until water uptake and cooking time were measured. At the Scottsbluff and Mitchell Ag Labs, each plot consisted of two 3.7-m rows with 1.11-m row width, with a sowing density of around 16 seeds per linear meter. Rainfall was supplemented with an overhead and furrow irrigation applied 10 times throughout the growing season with 412.0 mm (16.24 inches) in total in 2018 and 10 times with 409.5 mm (16.12 inches) in total in 2019. At maturity, all bean plants in the center 3.0-m section of the two rows were hand-pulled, and the seeds were harvested by using an Almaco plot thresher. Harvested seeds were dried at room temperature, cleaned to remove chaff, stones, and plant debris by using a seed cleaner (Clipper Office Tester, AT Ferrell Company Inc., Bluffton IN, USA), hand-cleaned to remove gravel and damaged or foreign seeds, and stored at a room-temperature until water uptake and cooking time were measured.

Water uptake and cooking time measurement

Seed moisture was equilibrated in a cold storage room at 4°C and 75% relative humidity and confirmed to be in the range of 10-14% before the bean samples were subjected to water uptake

and cooking time measurement. An automated Mattson cooker system was used to measure cooking times according to the method described by Cichy et al. (2015b). Briefly, 30 seeds were soaked in distilled water for 12 hours for genotypes grown in MI and 16 hours for genotypes grown in NE, and the soaked weight was measured after blotting soaked seeds with a paper towel. The water uptake was calculated as water absorbed by the seeds (g) {dry seed weight before soaking (g)}$^{-1}$ x 100. Twenty-five soaked seeds were placed on an automated Mattson cooker apparatus (Wang & Daun, 2005), cooked in boiling distilled water, and the 80% (20th) pin drop time was recorded as the cooking time. The cooking time of each field replication was measured. In Michigan, 9 lines in 2018 and 2 lines in 2019 had only one field replication planted or harvested. For these lines, two samples from the harvested plot were taken and treated as technical replications. The mean of the two field or technical replications were used as the cooking time for each environment and year.

Statistical analyses

The water uptake and cooking time data for MI and NE for 2018 and 2019 were transformed according to Box-Cox transformation (Box & Cox, 1964) by using the MASS package (Venables & Ripley, 2002) in R (R Core Team, 2017). The statistical analyses of the phenotypes collected on all trials were performed with the following mixed liner model: $Y_{ijkl} = \mu + G_i + E_j + Y_k + EY_{jk} + GE_{ij} + GY_{ik} + GEY_{ijk} + B(EY)_{jkl} + \varepsilon_{ijkl}$, where: Y_{ijkl} is the phenotypic value of the *i*th YBC genotype grown in the *l*th block of the *j*th environment in the *k*th year, μ is the grand mean, and G_i is a random effect of the *i*th genotype. E_j and Y_k are fixed effects of the *j*th environment and the *k*th year, respectively, EY_{jk}, GE_{ij}, and GY_{ik} are two-way interaction terms, GEY_{ijk} is a three-way interaction term, $B(EY)_{jkl}$ is a random effect of the *l*th block nested in the *j*th

environment and in the *k*th year, and ε_{ijkl} is an error term. The variance components of the model were used to calculate the broad-sense heritability of the traits (Fehr, 1987). The variance components were also used to partition the interaction terms into two parts, V and L. V represents variation due to heterogeneity of genotypes among various environments, and L represents the lack of correlation between genotypes among various environments (Cooper & DeLacy, 1994; Dickerson, 1962). If V accounts for a large portion of genotype x environment, selection would be more straightforward, and genotypes with the highest mean could be selected. If L is the more important portion of genotype x environment effect, the interaction would complicate selection due to changes in rank of genotypes in various environments. Only significant interactions of interest (i.e., genotype x environment and genotype x year) were examined for each trait.

The broad-sense heritability was also estimated within each environment by using the following mixed linear model: $Y_{ijk} = \mu + G_i + Y_j + GY_{ij} + B(Y)_{jk} + \varepsilon_{ijk}$, where Y_{ijk} is the phenotypic value of the *i*th genotype grown in *j*th year in the *k*th block, μ is the grand mean, G_i is the random effect of the *i*th genotype, Y_j is the fixed effect of the *j*th year, GY_{ij} is the interaction term of the *i*th genotype and the *j*th year, $B(Y)_{jk}$ is the random effect of the *k*th block nested in the *j*th year, and ε_{ijk} is the error term. The Best Linear Unbiased Estimators (BLUE: least squares means) were calculated for the 260 and 217 phenotyped individuals for cooking time grown in MI and NE, respectively by setting the genotype as a fixed effect and all other terms as random effects using the emmeans package (Lenth, 2021) in R for genome-wide association (GWA). All models were fitted using the lme4 (Bates et al., 2015) and the lmerTest (Kuznetsova et al., 2017) packages in R.

Genotyping of the YBC accessions and genetic identity controls

Five seeds of each YBC line were planted in the Michigan State University research greenhouse. Five trifoliate leaves per line were collected, frozen at −80°C, lyophilized, and milled by using a benchtop mill (Geno Grinder 2000, Spex Certiprep, Metuchen, NJ, USA). DNA was extracted from the milled leaves by using NucleoSpin Plant II Kit (Macherey-Nagel, Duren, Germany) following the 'Genomic DNA from plant' protocol. DNA was eluted in 50 μL of PE buffer. The purity of DNA was checked by running random samples of DNA on 1% agarose gels visualized using ethidium bromide stain. DNA concentration was measured by using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen), and 10 ng/μL of DNA was used for Genotyping-By-Sequencing (GBS) library preparation according to Elshire et al. (2011) with a single restriction enzyme, ApeKI. Barcoded DNA from 96 lines were pooled into one sequencing sample. Of the 96 adapter sequences, 11 were modified to increase specificity according to Thomas P. van Gurp (http://www.deenabio.com/services/gbs-adapters, accessed on April 20, 2018). The libraries were cleaned by using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) and sequenced with an Illumina standard HiSeq 4000, generating 150 bp single-end reads at RTSF Genomics Core at Michigan State University. Cutadapt ver. 1.16 was used to remove the adaptor sequence 'GAGATCGGAAGAG' with a minimum base quality score of 20, a minimum read length of 36, a maximum adapter sequence repeat of 3 times, and with the "remove flanking N bases from each read" option. FastQC ver.0.11.7 was used to confirm that there was no overrepresented sequences or adapter sequences in the libraries. An additional set of four control varieties, 1) Puebla 152 (black, MA race Mesoamerica), 2) Stampede (pinto, MA race Durango), 3) Montcalm, and 4) Talon (dark red kidneys, Andean race Nueva Granada) with known genepool

origins were sequenced by using the same method described above except that 50 bp single-end reads were generated.

SNP calling

SNP calling with the 18 control varieties and 295 YBC lines was carried out by using NGSEP software (Duitama et al., 2014). The libraries were de-multiplexed, and the reads of each line were aligned to the *P. vulgaris* genome ver.2.0 (Phytozome, https://phytozome.jgi.doe.gov/). Approximately 2 million SNPs and in-dels were called among the 295 YBC lines and 18 control varieties. SNPs were removed if they contained heterozygous calls, were non-biallelic, had a quality score < 40, were on scaffolds, or were on the predicted repetitive regions of the common bean genome (BeanWGS_RefV2_CIAT_UCDAVIS_repeatsNGSEPMergedMasked.txt, downloaded on December 18, 2018 from https://datadryad.org/stash/dataset/doi:10.5061/dryad.46pk7 (Lobaton et al., 2018)), after which 417,142 SNPs were retained. Heterozygous loci were removed because they are potentially genotyping errors as common bean is a self-pollinating species and is expected to have almost all homozygous genotypes. These 417,142-SNPs were the base SNP set, and it was filtered in different ways for the following uses.

Population diversity analyses

The base SNP set was further filtered for a minimum minor allele frequency (MAF) of 5% and < 20% missing genotyping data, which resulted in 2,234 SNPs. This SNP set was used for a principal component analysis using PLINK ver. 1.9 (Purcell et al., 2007). The first 10 principal components

were calculated, and the ggplot2 package (Wickham, 2016) in R (R Core Team, 2017) was used to visualize the diversity of the panel and the control varieties.

The base SNP set was filtered to remove SNPs in linkage disequilibrium (LD) to comply with the assumption by the STRUCTURE program that markers are at LD within subpopulations (Pritchard et al., 2000). Pairwise LD was calculated with a window size of 50 kb, a shift size of 5, and SNPs with $R^2 > 0.8$ were removed using PLINK ver. 1.9 (Purcell et al., 2007). The resulting 862 SNPs were used to conduct population structure analyses by STRUCTURE ver. 2.3.4 using the admixture model with independent allele frequencies and no prior population information for 10,000 burn-in periods followed by 50,000 Markov chain Monte Carlo (MCMC) replications (Pritchard et al., 2000). The first significant population split was determined by using delta K, which is a value related to the second order rate of change in the likelihood of the number of clusters (Evanno et al., 2005). Structure Harvester (Earl & VonHoldt, 2012) was used to determine delta K from the STRUCTURE output, which returned K=2 as the strongest population division. Common bean diversity studies have shown that there are two distinct genepools present in common bean: Andean and Middle American (Gepts et al., 1986; Gepts & Bliss, 1985; Koenig & Gepts, 1989). The control lines of known genepool origin were specified in the second run of STRUCTURE analyses with subpopulation numbers (K) from 2 to 4 for 10,000 burn-ins and 50,000 MCMC replications using prior genepool information of the 15 control varieties to assign individuals to clusters (Pritchard et al., 2009).

The base SNP set was filtered by using SNPhylo ver. 20140701 (Lee et al., 2014) with the default filtering conditions to remove SNPs with > 20% missing data, < 5% MAF, and/or pairwise LD of $R^2 > 0.8$ with a sliding window size of 500 kb. As a result, 965 SNPs were retained and used to construct a phylogenetic tree. The phylogenetic tree generated by SNPhylo was visualized by

using FigTree ver. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/, downloaded on December 16, 2020).

Genome-wide association study

The base 417,142-SNP set was further filtered to include the 260 and 217 YBC lines grown and phenotyped for water uptake and cooking time in MI and NE, respectively, and to exclude SNPs with > 20% missing data or < 5% MAF. As a result, 2,285 and 2,317 SNPs for MI and NE, respectively, were selected for GWAS. GWAS for cooking time in MI and NE were conducted by using the GAPIT package in R (Lipka et al., 2012) using the least squares means, which were adjusted for the year effects. Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) model was then used to detect associations between the SNPs and the traits (Huang et al., 2019). BLINK runs two fixed effect models: the first one tests associations between markers and trait values with pseudo quantitative trait nucleotides (previously associated SNPs with the trait; QTNs) as covariates and the second one optimizes the selection of QTNs to be included as covariates in the first model using Bayesian Information Criteria. BLINK has higher statistical power than its preceding model, Farm-CPU. A false discovery rate at $\alpha=0.05$ was used to call associations significant. The CMplot package (Yin, 2020) in R was used to generate QQ plots and Manhattan plots.

Results

Genetic diversity

The YBC is comprised of 295 *P. vulgaris* accessions with yellow-colored seeds primarily represented by major bean production regions of the world such as Africa (89 accessions),

followed by North America (83), South America (51), Europe (35), Central America and the Caribbean (32), and Middle East and East Asia (5) (**Table 1.1**). The color types of the YBC accessions included Amarillo light and dark, beige, brown, Canario (Mexican), Manteca, Mayocoba, and green-yellow (Njano) (**Figure 1.1**).

The genetic diversity of the YBC was evaluated through principal component analysis (PCA), analysis with STRUCTURE, and the generation of a phylogenetic tree. With PCA, most of the variance was explained by the first three principal components, with PC1 explaining 63.8%, PC2 explaining 16.1 % and PC3 explaining 4.5 % of the variation (**Figure S1.8**). PC1 separated the Andean and Middle American YBC lines, and PC2 separated the Middle Americans into two groups: defined by Mesoamerica and Jalisco-Durango races (**Figure 1.2A**). The accessions scattered in the middle of the plot were considered as admixtures between genepools. Many of the accessions in the Jalisco-Durango cluster are from North America, specifically Mexico, while the accessions in the Mesoamerica group are from Central America and Caribbean, mainly Haiti. PC3 separated Andeans into subgroups; The Andean lines from Africa and South America had higher values of PC3 while the Andean lines from Europe had lower values of PC3 (**Figure 1.2B**).

The STRUCTURE analysis at K=2 classified the YBC lines and the non-yellow 18 control genotypes into two clusters: Andean and Middle American similar to PC1 (**Figure 1.3A**). The YBC lines were classified as Andeans if the proportion of K1 was 70% or higher, and as Middle Americans if the proportion of K2 was 70% or higher. The lines that had less than 70% of both K1 and K2 were classified as admixtures. With the 70% cutoff, 203 YBC lines were classified as Andeans, 16 as admixtures, and 76 as Middle Americans. The membership proportions of the YBC in order of their geographical origin are shown in **Figure 1.3B**. K=2 shows that the accessions from Mexico contain both Andean and Middle American genotypes. At K=3, the Andeans were

separated into two subgroups; many of the U.S., Canadian, and CIAT lines from South America had high membership of K3. At K=4, both Andeans and Middle Americans were divided into two subgroups as expected from the results of K=2 and K=3.

The phylogenetic tree generated by SNPhylo is shown in **Figure S1.9**. As observed with PCA and STRUCTURE analyses, the YBCs were separated into two major groups: Andean and Middle American, and the lines that were considered as admixtures according to the population structure analysis at K=2 were placed between the two genepool clusters or with a longer distance from the Andean or Middle American clusters. Andeans formed a denser cluster than Middle Americans, which separated into Mesoamerica (leftmost) and Jalisco-Durango groups. In summary, the PCA, STRUCTURE, and SNPhylo analyses all agreed that the YBC contains Andean, Middle American, and admixed accessions.

Phenotypic diversity

The YBC was evaluated for plant growth habit (MI 2018), days to flower, seed size, water uptake, and cooking time (MI and NE in 2018 and 2019). The plant growth habit of the YBC by genepool is shown in **Table S1.8**. The Andean genotypes were 63% determinate and 37% indeterminate, whereas most of the admixtures and Middle American lines were indeterminate.

Days to flower

The YBC lines in each genepool exhibited variability for days to flower. Andeans had a higher proportion of genotypes that flowered early (about 37 days) than Middle Americans and admixture genotypes in all the environments and years (**Figure S1.10**). Out of 58 Middle American genotypes, 40 took 50 days or more to flower in MI 2018, but when almost the same set of genotypes were

grown in MI 2019, the most frequently observed days to flower of the Middle Americans was between 45-50. Many of the Jalisco-Durango lines from Mexico were unadapted in Michigan and Nebraska (i.e., did not flower or flowered/matured too late) due to photoperiod sensitivity. Days to flower measured in 2018 and 2019 were highly correlated within each environment (**Figure S1.11**).

100-seed weight

On average, the Andean YBC lines had the largest seed weights, the admixtures were intermeditate, and the Middle American genotypes were the lowest. This trend was observed in both environments and across years (**Figure S1.10**). Andeans and admixed genotypes produced smaller seeds in NE than in MI in both years, while Middle American genotypes produced seeds of similar size in all the environments. Since the plots were irrigated in NE, there should not have been drought stress; however, the Andeans and admixed genotypes might have been heat-stressed, flowered early, and produced smaller seeds under the NE environment. In 2018 and 2019, the number of days with temperatures above 30°C were 35 and 32, respectively. Temperatures above 30°C cause excessive flower drop and abortion of fertilized ovules (Rainey & Griffiths, 2005). The seed weights measured in 2018 and 2019 were highly correlated within each environment (**Figure S1.12**).

Water uptake and cooking time

The YBC seeds exhibited a large variability in water uptake and cooking time (**Table 1.2**). The water uptake and the cooking times of the YBC ranged from 7% to 132% and from 15 to 210 min, respectively, and the mean and the median cooking times were larger for NE than MI in both years.

Beans typically double in weight after being fully hydrated, which is equal to 100% water uptake. The YBC lines grown in MI showed normal distributions around 90-100% for water uptake in both years with a small number of lowly hydrating lines (**Figure 1.4A**). In contrast, the YBC grown in NE showed much flatter distributions with the medians lower than those grown in MI. The distributions of cooking time were similar to those of water uptake except that the cooking time distribution for NE 2018 was similar to a normal distribution but skewed towards longer cooking times (**Figure 1.4B**). In NE, the lines that had low water uptake had longer cooking times, whereas in MI, there was not a strong relationship between water uptake and cooking time (**Figure S1.13**).

The cooking times measured in each environment were moderately consistent across the two years, although genotypes grown in NE showed higher year-to-year variability (**Figure S1.14**). The cooking times of the genotypes grown in MI and NE were not highly correlated for either 2018 or 2019 (**Figure S1.15**). The cooking time distribution by genepool is shown in **Figure 1.5**. In MI, there was no obvious difference by genepool in the cooking time distributions, but the Middle Americans had a peak at a slightly shorter time in 2019. In NE, Andeans tended to be the faster cooking than admixtures and Middle Americans. There was no large difference in cooking time between seed size categories in MI for both years, whereas small-seeded beans took longer to cook in NE, especially in 2019 (**Figure S1.16**), and this is considered as a result of the lower water uptake by small-seeded beans, mainly for the Middle Americans (**Figure S1.17**). The genotype effects were only significant for cooking time while the year, environment, genotype x environment, environment x year, and genotype x environment x year were all significant for water uptake and cooking time (**Table 1.3**). The lack of significance for the genotype effect for water uptake may be related with the flat distribution of the phenotypes observed in NE in both years (**Figure 1.4A**). For water uptake, the genotype x environment was partitioned into 47% V and 53%

L. For cooking time, genotype x environment was partitioned into 1% V and 99% L (**Table 1.3**). The higher L for cooking time corresponded with the low correlation of cooking time among genotypes between the two environments (**Figure S1.15**). The broad sense heritability estimate of cooking time was low (0.34) when both environments were combined but was higher when estimated for each environment (0.80 for MI and 0.63 for NE, **Table 1.3**).

Table 1.4 shows the five fastest cooking beans grown in MI and NE. Four of the five fastest cooking beans in MI were classified as Middle Americans, and none were part of the Andean Diversity Panel. They were small-seeded, and their seed colors varied from beige to dark orange. In NE, all the five fastest cooking beans were Andean and had similar seed coat colors of lemon-yellow or yellow but differed in seed size ranging from 19.8 to 45.3 g per 100 seeds. YBC231 was the fastest cooking in both environments, and the other four fastest cooking beans in NE were also fast cooking (28 min or less) in MI.

GWAS for water uptake and cooking time

The QQ plots and Manhattan plots generated from the GWAS analyses for water uptake and cooking time are shown in **Figures 1.6 and 1.7**. For water uptake, significant SNPs were identified on Pv01, Pv02, Pv04, Pv08, and Pv09 (**Table 1.5**). For cooking time, significant SNPs were identified on Pv01, Pv03, Pv04, Pv06, Pv08, and Pv10 (**Table 1.5**). The phenotypic effect was generally larger for SNPs detected in NE than in MI, which corresponds to the wider distribution of both water uptake and cooking time in NE (**Figures 1.4A, 1.4B**). There was no overlap of significant SNPs between MI and NE for either water uptake or cooking time, thus associations were environment dependent. However, there were some overlaps for SNPs for both traits within each environment. The two significant SNPs detected on Pv08 for water uptake and cooking time

in MI were 677 kb apart. The SNP Chr04pos4548944 identified in NE was significant for both traits: water uptake and cooking time.

The major allele groups had shorter cooking times than the minor allele groups for all the significant SNPs in NE. For MI, the results were mixed. Two out of four significant SNPs had the major allele group having shorter cooking times (**Table 1.5**). These findings lend some degree of support to our original idea that yellow beans are a source of fast cooking time, with the growing environment playing an important part. The cooking time ranges of the five fast-cooking control varieties were 18-29 min in MI and 26-83 min in NE, and those of the three slow-cooking control varieties were 22-41 min in MI and 34-117 min in NE. The cooking time ranges of the YBC genotypes were larger than those of the control varieties, and some YBC genotypes cooked faster than the fast-cooking control varieties in each combination of environment and year, indicating a rich genetic diversity in cooking time of the YBC.

Discussion

Prior studies on yellow beans for consumer-focused traits such as fast cooking times and high iron bioavailability have produced interesting findings (Cichy, Wiesinger, et al., 2015; J. Wiesinger et al., 2018; J. A. Wiesinger et al., 2016). Given the diversity in shape, size, and colors within yellow beans and their importance in markets across the world, the YBC was assembled to facilitate the improvement and use of yellow beans. Genotypic and phenotypic characterization of the YBC for important traits will help assess the potential of the panel as a resource for breeding and genetic studies.

The YBC lines were genotyped by using the GBS technology and the genetic diversity was assessed using PCA, STRUCTURE, and phylogenetic analyses. All the analyses indicated that the

YBC lines can be grouped as Andean, Middle American, or admixture lines. The first major split of the panel separated Andean from Middle American accessions and was consistent with previous studies (Cichy, Porch, et al., 2015; Díaz & Blair, 2006; Oladzad et al., 2019). The existence of yellow beans in both genepools is advantageous in providing beneficial traits to other market classes because of the difficulty in inter-genepool hybridization in common bean (Gepts and Bliss, 1985).

Sixty-two genotypes in the YBC were classified as Mayocoba type, 32 of which were from South America. Among the 32 genotypes from South America, 27 were from CIAT, Colombia. Most of the accessions originating from South America were in the Andean cluster, which is expected considering that the Andean genepool includes accessions from this region. The Mayocoba type also included 19 U.S. breeding lines and one variety (YBC116, Myasi). Our STRUCTURE analysis showed that one of the Andean subgroups at K=3 consisted of genotypes from U.S. breeding lines and CIAT lines (**Figure 1.2B**). K3 in this case could be separating race Peru from other Andeans races because many of these lines were collected in Peru, Ecuador, Chile, or Colombia, where race Peru beans are known to exist, especially as yellow beans.

The Mesoamerica group contained many lines from Central America and the Caribbean (**Figure 1.2A**), and 27 of them were from Haiti. Haitian growers seemed to prefer Middle American landraces because of their high agronomic yield (Mainviel, 2019). Black beans are the most important seed type for Haitian consumers, but yellow beans are also consumed (Beaver et al., 2012). To our knowledge, this is the first study to report the genotypic and phenotypic diversity of these yellow beans from Haiti. Half the Mexican accessions were Middle American, and the other half were Andean when classified using STRUCTURE (**Figure 1.3B**). This highlights the diversity of yellow beans present in Mexico: Mexico is the center of origin for Middle American

germplasm, but local consumers have accepted and grow Andean yellow beans that must have been introduced originally from South America (Voysest, 2012).

The YBC accessions from Africa were found in all the three clusters in the PCA, indicating wide genetic diversity of African germplasm (**Figure 1.2A**). Similarly, whole-genome sequencing of selected beans and related species found that African accessions distributed into various races and genepool groups. The diversity of African germplasm likely resulted from the importation of germplasm developed in other parts of the world and the relatively lenient market class specification requirements (Lobaton et al., 2018).

The YBC showed 4.7- to 9.5-fold differences in cooking time, and the differences were more pronounced in NE (**Table 1.2**). The fast-cooking beans in MI were not necessarily fast-cooking in NE in this study (**Figure S1.15**), indicating strong environmental effects. Some studies reported limited environmental effects on cooking time (Cichy et al., 2019; Katuuramu et al., 2020), while a two-year study using a biparental RIL population grown in contrasting environments showed a large environmental effect on cooking time (Berry et al., 2020). Likewise, the strong environmental effect found in this study is probably due to the two contrasting environments: MI with high humidity and NE with low humidity. Indeed, the heritability estimates for those traits were higher when estimated for each environment (**Table 1.3**) and were consistent with previous studies which reported moderate to high heritability (A. Bassett, Kamfwa, et al., 2021; Cichy et al., 2019; Elia et al., 1997; Katuuramu et al., 2020). In addition, the beans were cooked in the production environment rather than being shipped to MI unlike some of the previous studies (Berry et al., 2020; Cichy et al., 2019), so environmental differences in handling and storage conditions may have been a part of the environmental effect. The NE production environment has the tendency to induce hardshell more so than the MI environment. Hardshell occurs in environments

where the weather is dry during seed filling (Castellanos et al., 1995). Seeds with hardshell do not absorb water during soaking and take longer times to cook, which is a hydrothermal process. Hardshell can increase variability within genotypes because it does not necessarily impact all seeds of a genotype but just a few. Overall, these two environments, MI and NE, are valuable for testing the effect of humid and arid growing conditions on cooking time.

For water uptake, over half of the genotype x environment was accounted for by L, the lack of correlation between genotypes in different environments. For cooking time, the L was essentially the sole source of genotype x environment interaction (**Table 1.3**), which is in accordance with the results observed in **Figure S1.15**. Since high L impedes gain from selection due to changes in rank of genotypes among environments, these results suggest that selection in specific environments or testing in a large number of environments may be necessary to select high-performing lines for these traits. However, this estimate is based on only two environments; a larger number of environments should be used in order to estimate more generalizable V and L. Previous studies with large numbers of environments (9-15) have concluded that the genotype x environment interaction effect on water uptake and cooking time is small and therefore a small number of environments are needed for evaluation of those traits (Cichy et al., 2019; Katuuramu et al., 2020). It would be more practical for breeding programs to estimate V and L with this type of multi-environment study.

This study identified fast cooking yellow beans including Manteca and other types that cooked in 15-31 min in MI and NE (**Table 1.4**). In addition, 15 more YBC lines grown in MI cooked within 20 min. The cooking times of promising lines should be evaluated in the target production environments because YBC281 (SSIN 533 from CIAT-Uganda) and YBC192 (Roba-1, CIAT-Uganda) were two of the five fastest cooking in MI but took 80 and 97 min, respectively,

in NE. The YBC lines that showed constantly short cooking times in both environments will be valuable genetic resources for developing fast-cooking bean varieties with stable performance in different environments. Some Manteca beans were reported to be fast-cooking (Cichy et al., 2019; J. Wiesinger et al., 2018; J. A. Wiesinger et al., 2016). Tannin content and cooking times were positively correlated (Elia et al., 1997), and the lack of polyphenols such as proanthocyanidins (condensed tannins) in Manteca beans (Beninger et al., 1998) may be one of the reasons for fast cooking times of Manteca beans. It is interesting that the YBC contained some fast-cooking beans that were not Manteca type. The YBC lines were held in ambient conditions and cooked relatively soon after harvest. Thus, tannins present in the seeds may not have impacted cooking times as much as they may have if beans were stored longer under unfavorable conditions, which can induce tannin polymerization leading to long cooking times (Martín-Cabrejas et al., 1997; Stanley, 1992). In addition, there seems to be more mechanisms for the short cooking times as varying polyphenol profiles have been reported in some yellow beans (Hart et al., 2020). Beans in MI were stored in cold storage until cooking time measurement, and this may be another reason for the differences in the top 5 fastest cooking beans in MI and NE. The cooking times in MI were likely the shortest possible cooking time of the genotypes because the seeds were stored in cold storage, cooking time was measured within 4-6 months after harvest, and the moisture was equilibrated to 10-14% before soaking. Consumers are expected to experience longer cooking times than the potential of genotypes because beans are typically stored under room temperature in the supply chain and could be 1-2 years old before they are purchased and cooked. More attention should be paid to age and storage conditions of beans in order to deliver the benefit of short-cooking beans to consumers.

All the significant SNPs for water uptake were novel when compared with previous studies using markers with known physical positions (A. Bassett, Kamfwa, et al., 2021; A. Bassett,

Katuuramu, et al., 2021; Berry et al., 2020; Cichy, Wiesinger, et al., 2015; Diaz et al., 2021; Soltani et al., 2021). Soltani et al. (2021) found QTL for water uptake between 43.1–53.4 Mbp on Pv03 by a bulked segregant analysis (BSA) with a population derived from a cross between two red kidney beans with contrasting water imbibition rates. Then they narrowed it down to 51,426,054 - 51,544,057 bp region by whole genome sequencing, where they found a candidate gene. However, no SNPs on Pv03 were associated with water uptake in this study. This region was not a genotyping gap because there was a SNP in this region, Chr03pos51541190, but it was not associated with either cooking time or water uptake. However, one of the significant SNPs for cooking time in NE, Chr03pos50288381, was 1.1 Mbp away from the narrowed-down region. This SNP associated with cooking time might be related to water uptake because cooking time and water uptake were highly correlated in NE ($R = -0.72$ for 2018 and $R = -0.80$ for 2019; $p<0.0001$ for both years).

Several SNPs on Pv02 and Pv04 were significant in this study, and QTLs were detected on the same chromosomes in some previous studies with no physical position available (Pérez-Vega et al., 2010; Sandhu et al., 2018). One of the significant SNPs for water uptake on Pv08 in MI, Chr08pos1429025, had a mean water uptake that was 14 percentage points higher than the minor allele group (**Table 1.5**). Since the minor allele frequency is 0.06 for this SNP, the major allele with higher mean water uptake of this SNP is expected to be found more commonly. Therefore, this SNP could be useful in filtering out genotypes with lower water uptake potential. For all the SNPs associated with water uptake in NE, the allele groups that had higher water uptake contained only Andean or admixtures, except for Chr02pos49223533, of which the minor allele (higher water uptake) group had one Middle American and 27 Andean or admixtures. Andeans having the higher water uptake alleles corresponded with the observation that larger seeded beans had higher water uptake in NE (**Figure S1.17**).

Some of the SNPs associated with cooking time were supported by previous studies. The SNP on Pv06, Chr06pos27715283, in MI was within a QTL found by Berry et al. (2020): CT.6.1, 24,511,850 - 27,381,730 bp on Pv06, was previously detected at two locations in Tanzania. The authors used a RIL population derived from a cross of two brown beans with contrasting cooking times. The significant SNP on Pv08, Chr08pos62054742 in MI and S08_62659170, the SNP associated with cooking time in the Andean Diversity Panel reported by Bassett et al. (2021a), were close (604 kb apart) and were within the range of QTL, CT.8.2, for cooking time detected in a yellow bean RIL population (A. Bassett, Katuuramu, et al., 2021). The QTL study used a yellow bean RIL population generated by a cross between a fast-cooking Manteca type bean (Ervilha, YBC129) and a slow cooking Njano type bean (PI527538, YBC130). In this study, the major allele group of Chr08pos62054742 had 3 minutes shorter mean cooking time than the minor allele group (**Table 1.5**), and YBC129 was in the minor allele (long-cooking) group. This was consistent with the QTL on Pv08, where the fast-cooking YBC129 conferred the longer-cooking allele (A. Bassett, Katuuramu, et al., 2021). Overall, one QTL and three significantly associated SNPs for cooking time were found at 53.03 - 62.66 Mbp on Pv08 in three studies with Andean and/or yellow beans, suggesting that this region is worth further investigation for fast cooking time. The QTL and the SNPs could be informative markers for selection, especially yellow beans for cooking time.

The significant SNP on Pv10, Chr10pos43615359, in MI had the lowest *p*-values among all the SNPs associated with cooking time (**Table 1.5**). This SNP was in the range of a QTL, CT10.1 (37.83 - 43.84 Mb), detected with the highest LOD score among the QTLs for cooking time in the aforementioned yellow bean RIL population (A. Bassett, Katuuramu, et al., 2021). Considering the strong and consistent association between the two studies, this region on Pv10 will be another region to focus on to investigate the mechanism of cooking time of yellow beans.

Interestingly, YBC129 conferred the short cooking allele at CT10.1, while both YBC129 and YBC130 were in the major allele group in this study, which had a longer cooking time than the minor allele group at the SNP (**Table 1.5**). The *J* gene that controls non-darkening of seed coat resides in the same CT10.1 genomic region (Erfatpour et al., 2018; Erfatpour & Pauls, 2020). Post-harvest darkening is an important consumer-acceptance trait as darkened seeds are considered as old and long-cooking by consumers, leading to reduced economic values (Erfatpour et al., 2018; Felicetti et al., 2012). It is an important quality trait of yellow beans as some genotypes are susceptible to post-harvest darkening while others are not (unpublished data).

The significant SNP on Pv01, Chr01pos1528388, in NE was at a distal region of Pv01. Three QTLs for cooking time were previously detected on Pv01 (Garcia et al., 2012); however, their physical locations were unknown. On Pv04, the marker Chr04pos4548944 was significantly associated with water uptake and cooking time in NE. Chr04pos4548944 is located in the coding region of Phvul.004G038700, which encodes a polygalacturonase (EC 3.2.1.15). Cotyledon cell separation is a pivotal process that allows bean softening, and cell separation is preceded by pectin solubilization in the middle lamella (Chigwedere et al., 2019; Rockland & Jones, 1974). Pectins are complex polysaccharides made up of polygalacturonans as a backbone (Ridley et al., 2001). Polygalacturonase activities were increased during soaking and were associated with increased polygalacturonan extractability during cooking and decreased cooking time as compared to unsoaked beans (Martínez-Manrique et al., 2011). Therefore, a polymorphism in Phvul.004G038700 encoding a polygalacturonase may result in changes in the degree of polygalacturonan degradation in pectin, resulting in longer cooking times. The minor allele group for this SNP had 21.7 minutes longer cooking time and 21.7 percentage points lower water uptake than the major allele group (**Table 1.5**). This SNP was only significant in NE potentially because

water uptake may have been a limiting factor for pectin solubilization in partially hydrated seeds in NE, resulting in longer cooking times. Phvul.004G038700 had a 94.3% similarity score with Glyma.13G054600.1 and Glyma.19G032600.1, both of which encode polygalacturonase in soybean genome *G.max* Wm82.a2.v1 (Phytozome, ver.13 https://phytozome-next.jgi.doe.gov). Glyma.19G032600.1 was within the 0.7-7.3 Mbp region of Gm19 syntenic to Pv04, and Chr04pos4548944 was within the 0-77 Mbp region of Pv04 syntenic to Gm19 (McClean et al., 2010). Previous studies with soybean identified some QTL for water uptake traits on Gm05, Gm12 (Hirata et al., 2014), Gm07, Gm08, Gm15, Gm16, and Gm17 (Molnar et al., 2012; Ott et al., 2011), but these chromosomes did not have blocks syntenic to Pv04 (McClean et al., 2010), where Chr04pos454894 is located. A significantly associated SNP for cooking time with the ADP was also reported by Bassett et al. (2021a) 592 kbp upstream of this SNP.

Conclusions

This study evaluated the genotypic and phenotypic diversity of a yellow bean diversity panel in two environments in the U.S. for two years. The genotypic data for the YBC indicated that the YBC contains Andean, Middle American, and admixed individuals. A wide diversity was observed for growth habit, days to flower, seed weight, cooking time, and water uptake. Cooking times ranged from 15 to 210 min across all the environments and years, and the fastest cooking beans were of various yellow colors and sizes. The environmental effect on water uptake and cooking time was significant, and no SNPs associated with those traits overlapped between the two environments, indicating the importance of testing multiple production environments. SNPs on Pv04, Pv06, Pv08, and Pv10 significantly associated with cooking time were supported by previous studies, which calls for further investigation. A significant SNP on Pv04 was in the coding

region of polygalacturonase gene, and it was considered as a candidate gene for the fast-cooking trait. Considering the diversity and the significantly associated SNPs for cooking time, the YBC will serve as a valuable resource for bean breeding for consumer-focused qualities and for genetic studies to elucidate the mechanism of the fast-cooking trait.

Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture AFRI (Award Number #: 2017-67013-26212), Michigan State University, the National Science Foundation Research Traineeship Program (DGE-1828149) awarded to Rie Sadohara, and the Nebraska Hatch Project (NEB43-116). The authors thank Anna Akariza, Amber Bassett, Gasana Ingabiregasana, Miranda Haus, Sharon Hooper, Queen Iribagiza, Hannah Jeffery, Hannah Peplinski, Scott Shaw, Jason Wiesinger, and Evan Wright for assistance with cooking, collecting plant tissues, and field operations. The authors express gratitude to Dr. Phil McClean at North Dakota State University for his advice in analyzing the genetic diversity of the YBC.

APPENDICES

APPENDIX A:

CHAPTER 1 TABLES AND FIGURES

Table 1.1 Origin of the Yellow Bean Collection entries.

Region	Number of entries	
Africa		89
Uganda	36	
Tanzania	28	
Angola collection	11	
Kenya	5	
Burundi	3	
Other	3	
Unknown	3	
South America		51
Colombia	38	
Ecuador	5	
Brazil	3	
Peru	3	
Bolivia	1	
Chile	1	
Central America & the Caribbean		32
Haiti	27	
University of Puerto Rico	3	
Caribbean collection	1	
Nicaragua	1	

Region	Number of entries	
North America		83
Mexico	41	
United States	35	
Canada	7	
Europe		35
Netherlands	11	
Bulgaria	5	
Hungary	6	
Ukraine	3	
Macedonia	2	
Georgia	1	
Other	5	
Unknown	2	
Middle East		3
Iran	2	
Turkey	1	
East Asia		2
India	2	
Total		295

Table 1.2 The range, mean, and median of water uptake and cooking times of the YBC grown in MI and NE in 2018 and 2019.

	--------- Water uptake (%) ---------				--------- Cooking time (min) ---------			
Environment - Year	MI 2018	MI 2019	NE 2018	NE 2019	MI 2018	MI 2019	NE 2018	NE 2019
Range	45-114	51-108	16-125	7-132	15-68	16-80	22-130	22-210
Fold-difference	2.5	2.1	7.7	18.1	4.7	5.0	5.9	9.5
Median	99	94	85	54	26	27	38	71
Mean	98	93	82	59	26	28	44	75
SD	7	8	24	27	6	7	20	33

Table 1.3 The *p*-values of the factor effects and broad-sense heritability for water uptake and cooking time.

Factor	*p*-value	
	Water uptake	Cooking time
Genotype (G)	0.12	0.0043
Year (Y)	<0.0001	<0.0001
Environment (E)	<0.0001	<0.0001
G x E	<0.0001	<0.0001
V[a] (%)	47	1
L[b] (%)	53	99
G x Y	1.0	1.0
E x Y	<0.0001	<0.0001
G x E x Y	<0.0001	<0.0001
H^2 [c]- overall	0.18	0.34
H^2 - MI	0.76	0.80
H^2 - NE	0.77	0.63

[a]V: GxE component due to heterogeneity of genotypes among environments
[b]L: GxE component due to lack of correlation among genotypes among environments
[c]H^2: Broad-sense heritability

Table 1.4 The five fastest cooking YBC lines phenotyped in MI and NE. The cooking times and 100 seed weights are the means of the two years in each environment.

MI					
ID	YBC231	YBC281	YBC297	YBC192	YBC301
Genotype name	CAN 44	SSIN 533	F7500	Roba-1	M0181
Cooking time (min)	15.3	16.2	16.6	17.3	18.6
Part of ADP[a]	No	No	No	No	No
Country of Origin	Colombia	Uganda	Haiti	Uganda	Haiti
Source	CIAT	CIAT-Uganda	Farmer	CIAT-Uganda	Market
Genepool[b]	Andean	MA	MA	MA	MA
100 seed weight (g)	27.1	29.4	22.0	21.8	23.7
					

NE					
ID	YBC231	YBC071	YBC239	YBC129	YBC214
Genotype name	CAN 44	PI433611	CAN 79	Ervilha	NE-17-7
Cooking time (min)	23.0	23.5	27.0	30.3	30.5
Part of ADP[a]	No	No	No	Yes (ADP0512)	No
Country of Origin	Colombia	United States	Colombia	Angola	United States
Source	CIAT	US GRIN	CIAT	Angola collection	University of Nebraska
Genepool[b]	Andean	Andean	Andean	Andean	Andean
100 seed weight (g)	19.8	29.1	37.7	45.3	35.4
					

[a]ADP: Andean Diversity Panel; [b]Genepool is based on the STRUCTURE classification at K=2; [c]MA: Middle American.

Table 1.5 QTL identified by the genome-wide association analysis for water uptake and cooking time in environments MI and NE.

				----- Major allele -----			----- Minor allele -----				
SNP	Chr.	Position	FDR-adjusted *p*-values	Allele	n	Mean phenotypic value	Allele	n	Mean phenotypic value	MAF	SNP effect[a]
Water uptake (%)											
MI											
Chr02pos24852892	2	24,852,892	1.4E-04	AA	225	95.2	CC	19	97.7	0.08	−2.5
Chr04pos25006646	4	25,006,646	2.5E-07	TT	156	97.5	AA	62	91.3	0.28	6.1
Chr08pos1429025	8	1,429,025	3.8E-12	AA	236	96.4	GG	14	82.1	0.06	14.3
Chr08pos62731875	8	62,731,875	3.4E-07	AA	160	97.1	GG	75	91.4	0.32	5.7
NE											
Chr01pos45386764	1	45,386,764	4.5E-03	AA	166	75.5	GG	22	50.6	0.12	24.9
Chr02pos49223533	2	49,223,533	3.1E-02	GG	157	67.7	TT	28	84.8	0.15	−17.1
Chr04pos4548944	4	4,548,944	4.5E-03	CC	141	77.2	AA	37	55.5	0.21	21.7
Chr09pos35530354	9	35,530,354	2.4E-03	GG	194	68.5	AA	19	96.2	0.09	−27.7
Cooking time (minutes)											
MI											
Chr03pos3019420	3	3,019,420	1.2E-03	GG	190	26.0	AA	34	30.5	0.15	−4.5
Chr06pos27715283	6	27,715,283	1.1E-03	GG	222	27.0	AA	14	22.6	0.06	4.4
Chr08pos62054742	8	62,054,742	3.1E-02	GG	152	25.7	CC	68	28.7	0.31	−3.0
Chr10pos43615359	10	43,615,359	2.1E-10	CC	224	27.2	GG	25	21.8	0.10	5.5
NE											
Chr01pos1528388	1	1,528,388	2.8E-02	CC	178	51.8	AA	25	66.7	0.12	−14.9
Chr03pos50288381	3	50,288,381	4.9E-02	AA	181	49.7	CC	12	87.3	0.06	−37.6
Chr04pos4548944	4	4,548,944	3.5E-05	CC	141	47.8	AA	37	69.5	0.21	−21.7
Chr08pos511826	8	511,826	1.3E-04	AA	180	49.6	TT	19	83.4	0.10	−33.8
Chr10pos9440210	10	9,440,210	1.4E-05	AA	151	48.8	TT	24	78.3	0.14	−29.5

[a]SNP effect is the difference between the mean cooking times of the YBC lines with the major and minor alleles.

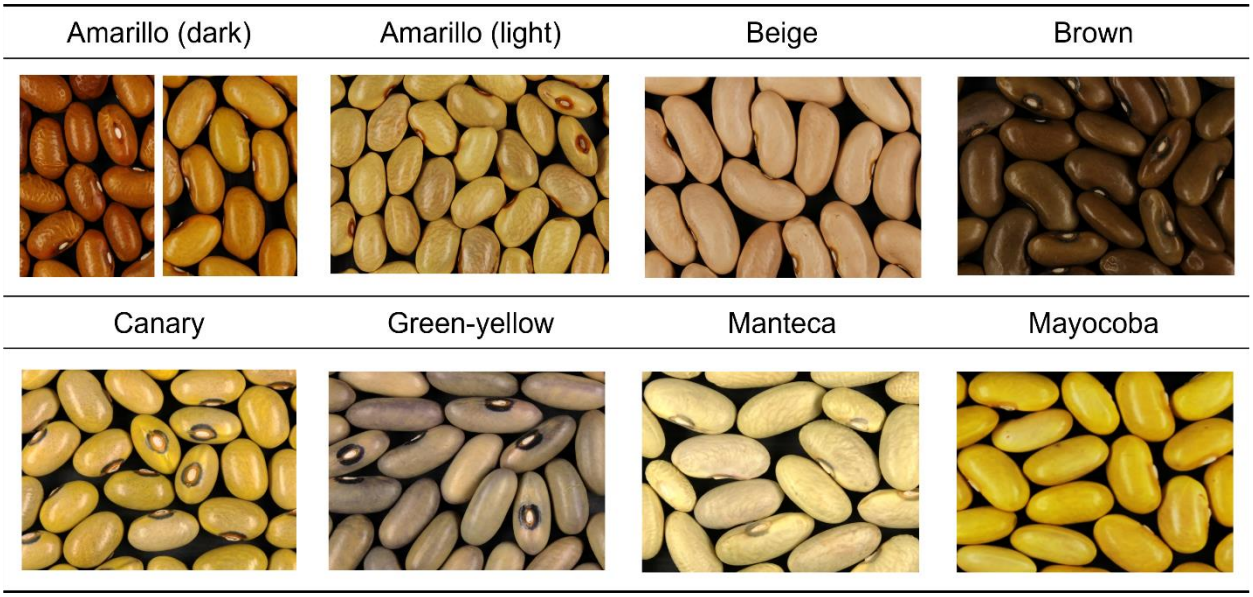
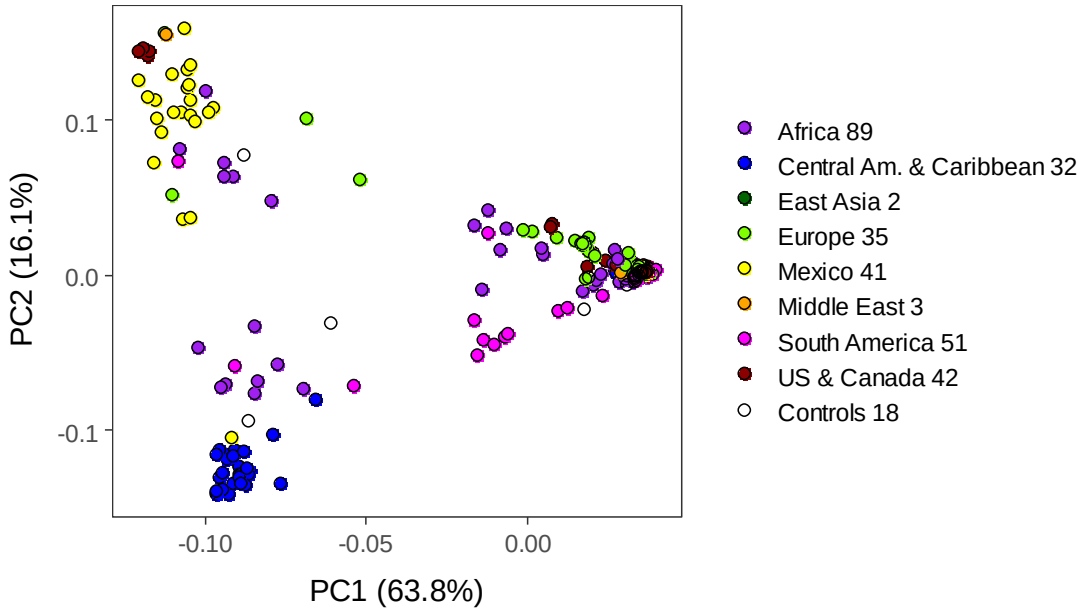


Figure 1.1 Examples of various shapes and colors of yellow beans.

A



B

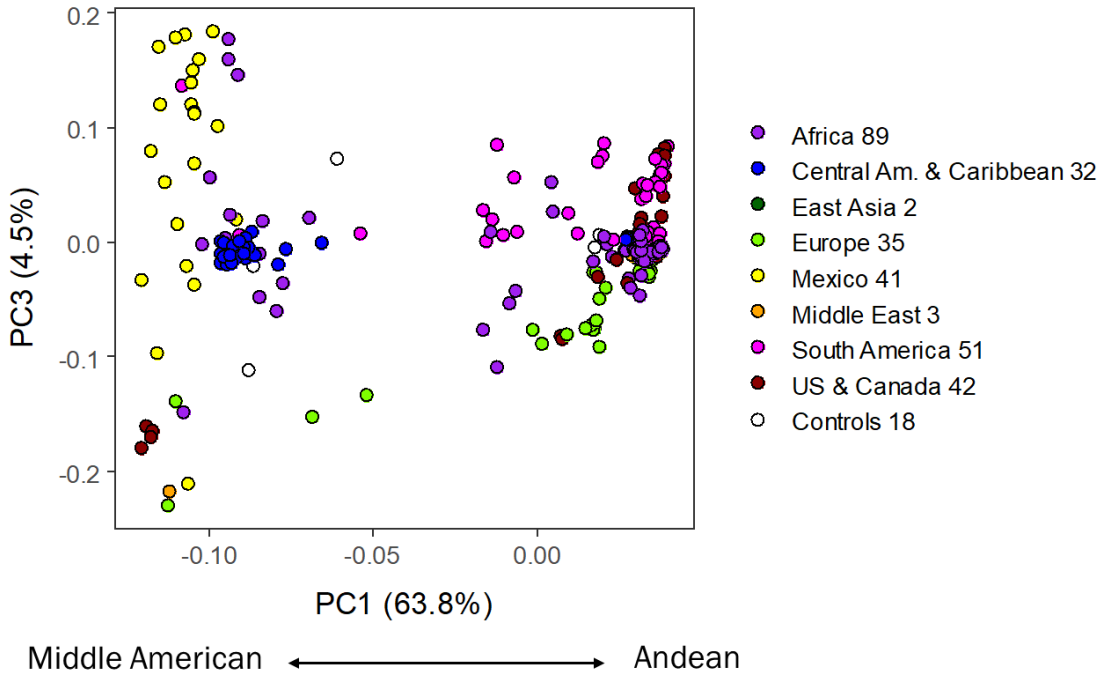
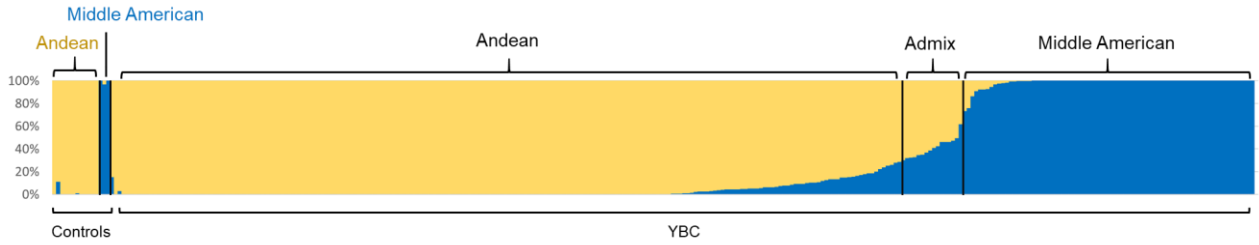


Figure 1.2 Biplots of principal component analysis of the YBC. **A**: PC1 and PC2 by the region of origin. The numbers in the legend indicate the number of accessions originated from each region. Arrows indicate varieties with known genepool or race origin.; **B**: PC1 and PC3 by the region of origin.

A



B

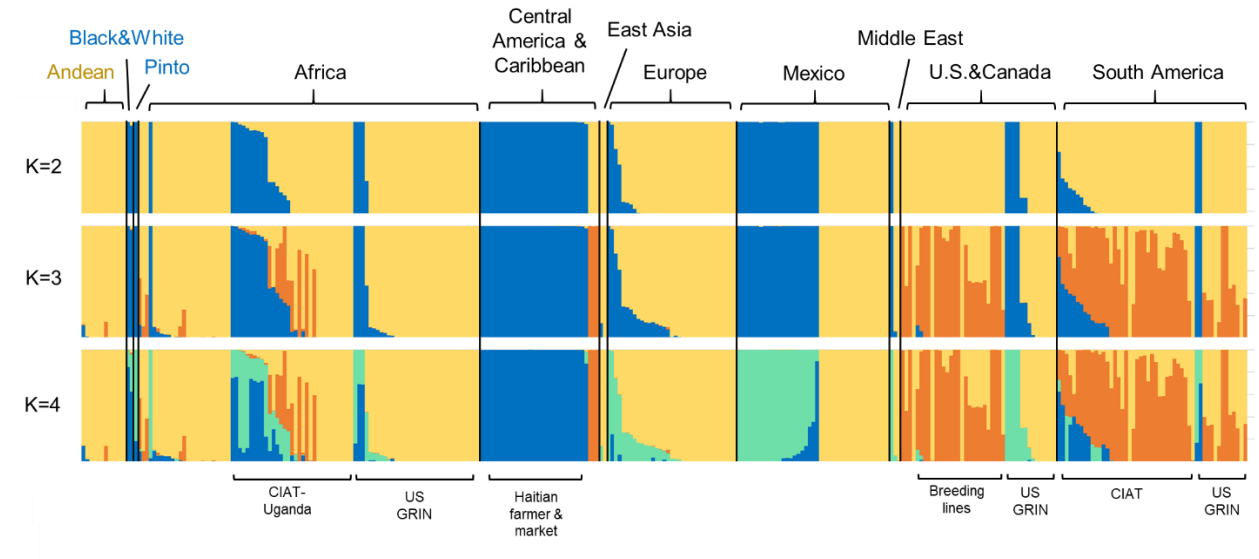


Figure 1.3 Population membership of the YBC lines and the controls determined by STRUCURE. **A**: Classification by the K1 membership (K=2); **B**: Classification by the geographical and institutional origin (K=2-4).

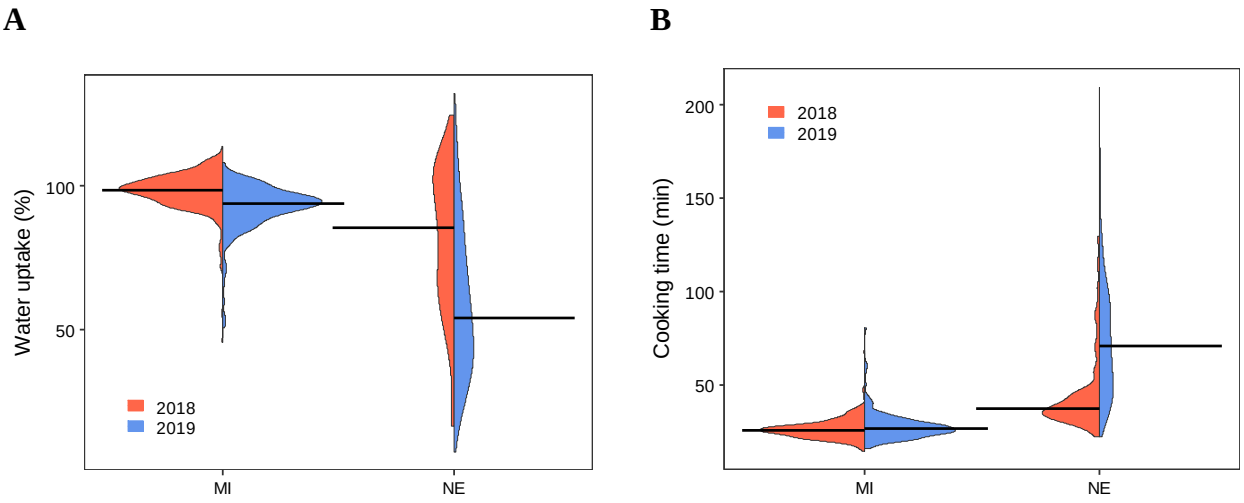


Figure 1.4 Phenotypic distribution of **A**: mean water uptake and **B**: mean cooking time of the YBC grown in MI and NE in 2018 and 2019. The black bars represent the median in each environment and year.

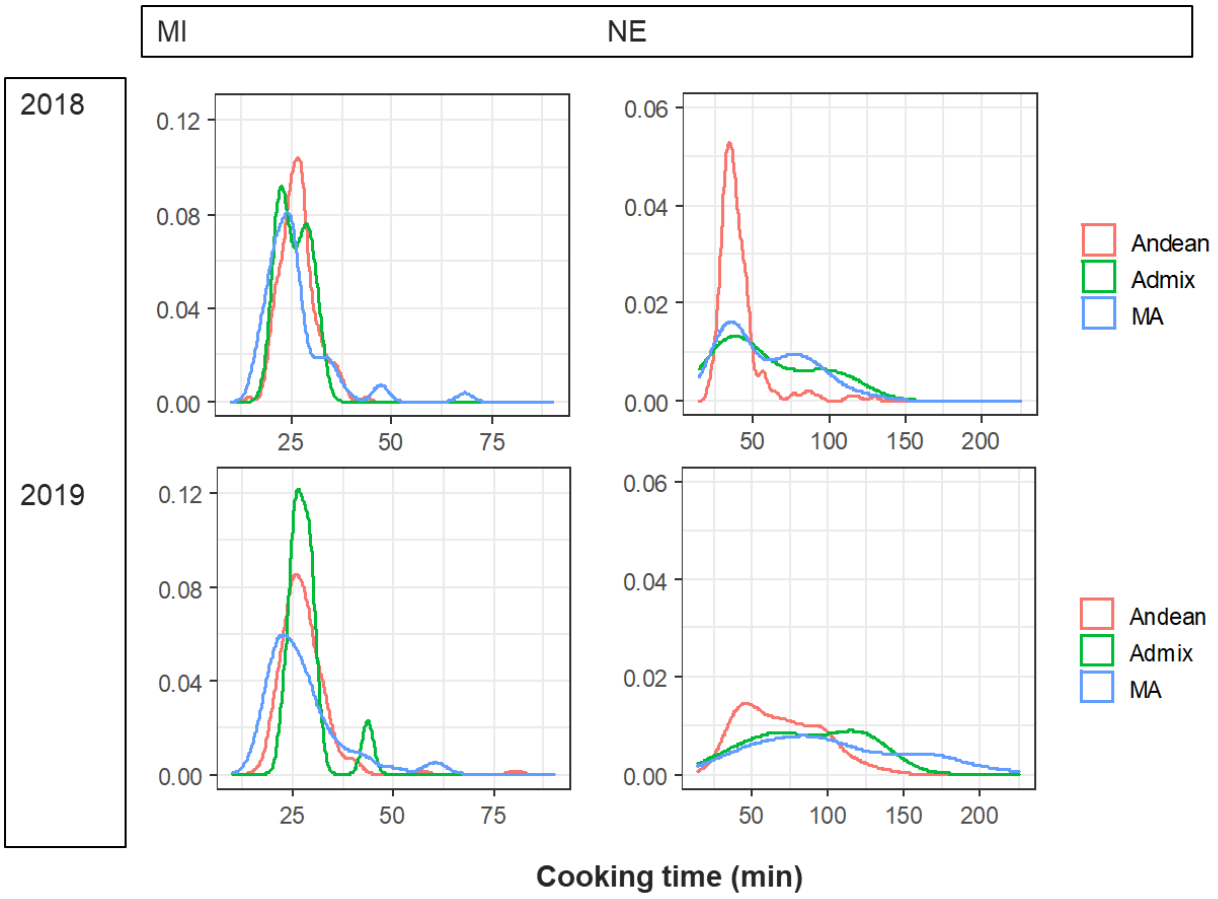


Figure 1.5 Density plots of cooking time by genepool of the YBC measured in MI and NE in 2018-2019. MA: Middle American.

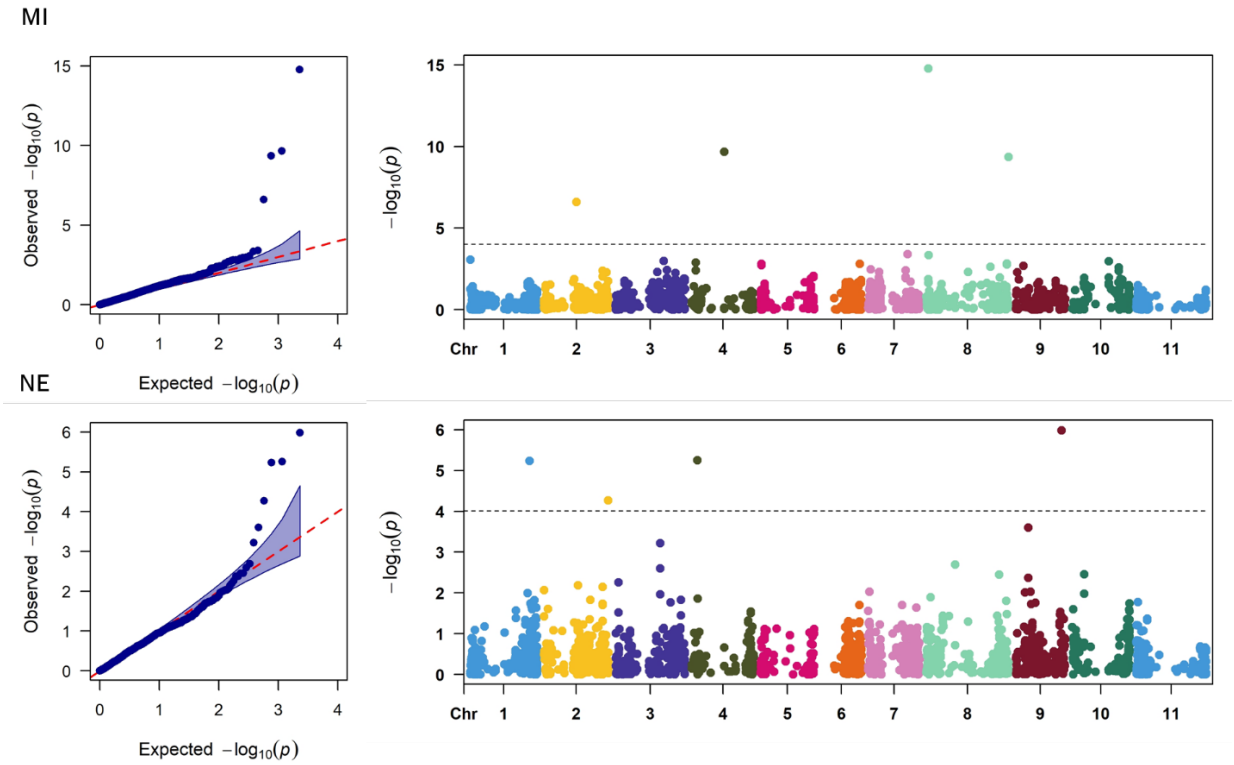


Figure 1.6 Genome-wide association analysis of water uptake. The QQ plots on the left show the model fit, and Manhattan plots show the *p*-values of SNPs for associations with cooking time in MI and in NE. The gray dotted lines indicate the false discovery rate-adjusted threshold for *p*-values (α=0.05).

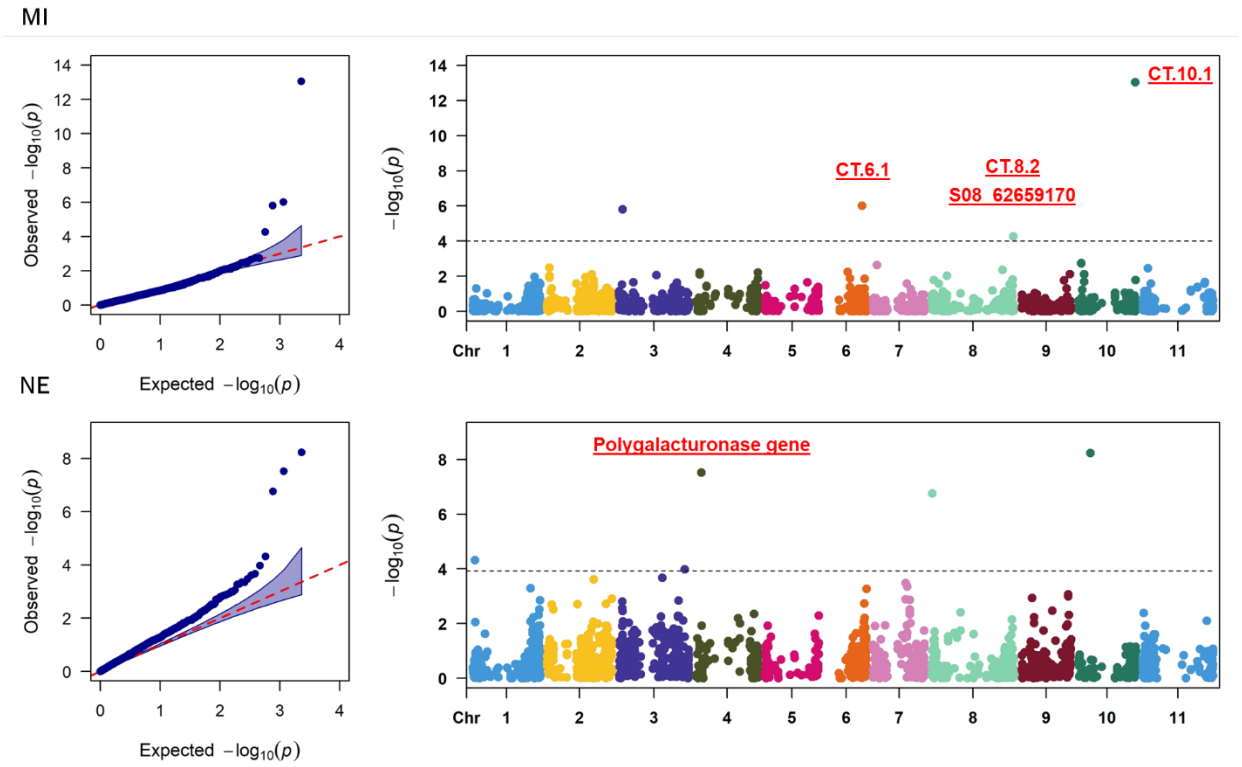


Figure 1.7 Genome-wide association analysis of cooking time. The QQ plots on the left show the model fit, and Manhattan plots show the *p*-values of SNPs for associations with cooking time in MI and in NE. The gray dotted lines indicate the false discovery rate-adjusted threshold for *p*-values (α=0.05). CT.6.1 and CT.8.2 are QTL reported by Bassett et al. (2021b). CT.10.1 is a QTL reported by Berry et al. (2020). S08_62659170 is a SNP significantly associated with cooking time (604 kb apart from the significant SNP in this study) (A. Bassett, Kamfwa, et al., 2021). The significant SNP on Pv04 detected in NE is in the coding region of a polygalacturonase gene (Phvul.004G038700).

APPENDIX B:

CHAPTER 1 SUPPLEMENTAL TABLES AND FIGURES

Table S1.6 The control varieties of the YBC.

YBC control ID	ID	Genotype name	Cooking time classification	Country of Origin	Source	Seed type	Genepool
Ctl001	Snowdon	Snowdon	-	United States	MSU[a]	White kidney	Andean
Ctl002	ADP0452[b]	INIAP425	Fast[c]	Ecuador	Ecuador breeding	White	Andean[d]
Ctl003	ADP0469	PI527521	Fast[c]	Burundi	USGRIN	White	Andean[d]
Ctl004	Merlin	Merlin	-	United States	MSU	Navy	MA[e]
Ctl005	Red Hawk	Red Hawk	-	United States	MSU	Dark Red Kidney	Andean
Ctl006	ADP0367	G23086	Fast[f]	Malawi	ADP	Cranberry	Andean[d]
Ctl007	ADP0434	PR0737-1	Slow[f]	Puerto Rico	ADP	Red mottled	Admix[d]
Ctl008	ADP0436	JB178	Fast[f]	Dominican Republic	ADP	Red mottled	Andean[d]
Ctl009	ADP0515	Katarina, Kibala	Slow[f]	Angola	ADP	Cranberry	Andean[d]
Ctl010	ADP0618	ACElk	Fast[f]	Canada	ADP	Light Red Kidney	Andean[d]
Ctl011	ADP0687	Pink Panther	Slow[f]	United States	ADP	Light Red Kidney	Andean[d]
Ctl012	ADP0680	Clouseau	Moderate[f]	United States	ADP	Light Red Kidney	Andean[d]
Ctl013	RWR2245	RWR2245	-	Rwanda	Katuuramu, Uganda	Red mottled	Unknown
Ctl014	RWR2154	RWR2154	-	Rwanda	Katuuramu, Uganda	Cranberry	Unknown

[a]MSU: Michigan State University; [b]ADP: Andean Diversity Panel; [c]: Cooking time classification by Cichy et al. (2015b); [d]Genepool information by Cichy et al. (2015a); [e]MA: Middle American; [f]: Cooking time classification by Cichy et al. (2019)

Table S1.7 The number of YBC lines phenotyped at each environment.

Trait	Environment and year	Andean	Admix	Middle American	Total
Days to flower	MI 2018	194	13	58	265
	MI 2019	192	12	56	260
	NE 2018	182	12	32	226
	NE 2019	175	9	30	214
Seed weight	MI 2018	191	12	57	260
	MI 2019	192	12	56	260
	NE 2018	180	10	31	221
	NE 2019	175	9	31	215
Water uptake, cooking time	MI 2018	191	12	57	260
	MI 2019	191	12	55	258
	NE 2018	177	10	30	217
	NE 2019	171	9	30	210

Table S1.8 The growth habit (MI, 2018) of the Andean, Admix, and Middle American groups in the YBC (determined at K=2 by STRUCTURE).

		Determinate		Indeterminate	
	(n)	(Count)	(%)	(Count)	(%)
Andean		120	63	70	37
Admix		3	23	10	77
Middle American		4	7	55	93
Total		127	-	135	-

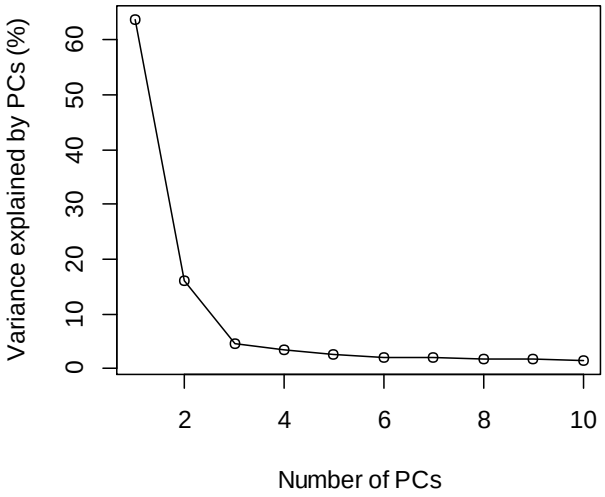


Figure S1.8 The variance of the YBC explained by the first 10 principal components.

Figure S1.9 A phylogenetic tree of the YBC. Genotypes in red and blue are Andean and Middle American control varieties, respectively, and genotypes in green are those considered to be admixtures by the STRUCTURE analysis at K=2.

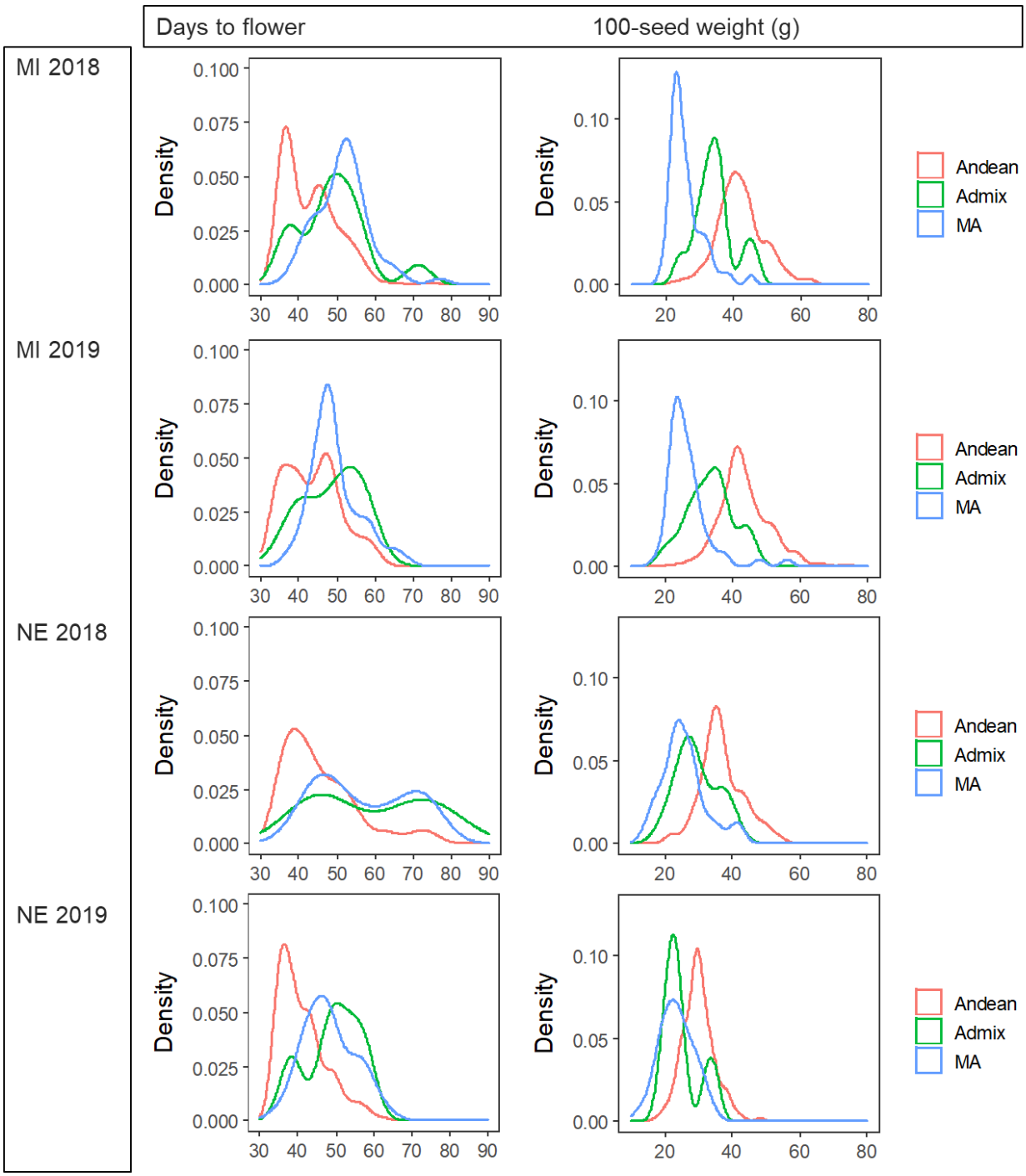


Figure S1.10 Density plots of days to flower and 100-seed weight by genepool of the YBC measured in MI and NE in 2018-2019. MA: Middle American.

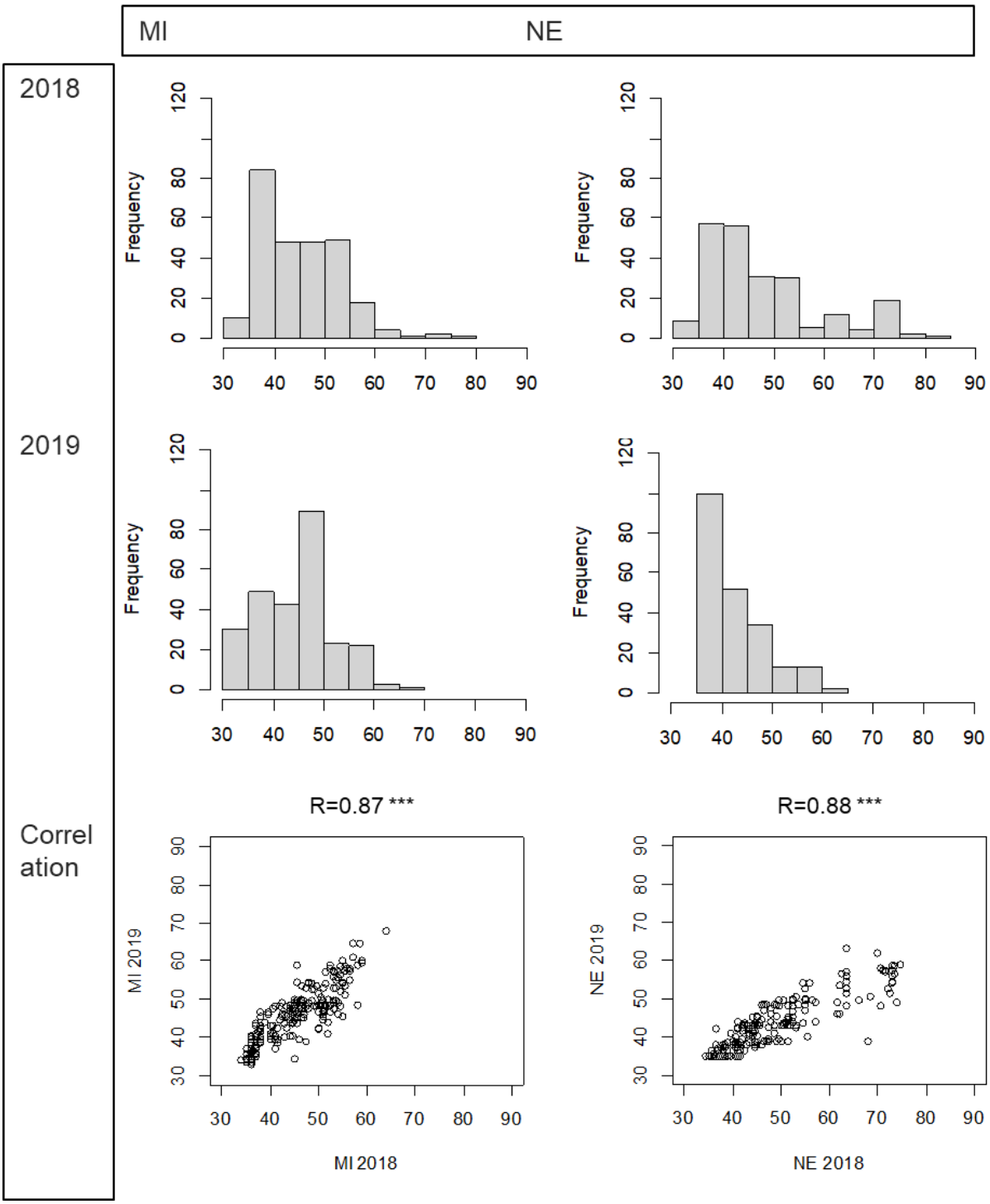


Figure S1.11 Phenotypic distribution of days to flower measured in MI and NE in 2018-2019 and the correlation of the 2 years within each environment. (***: p<0.001).

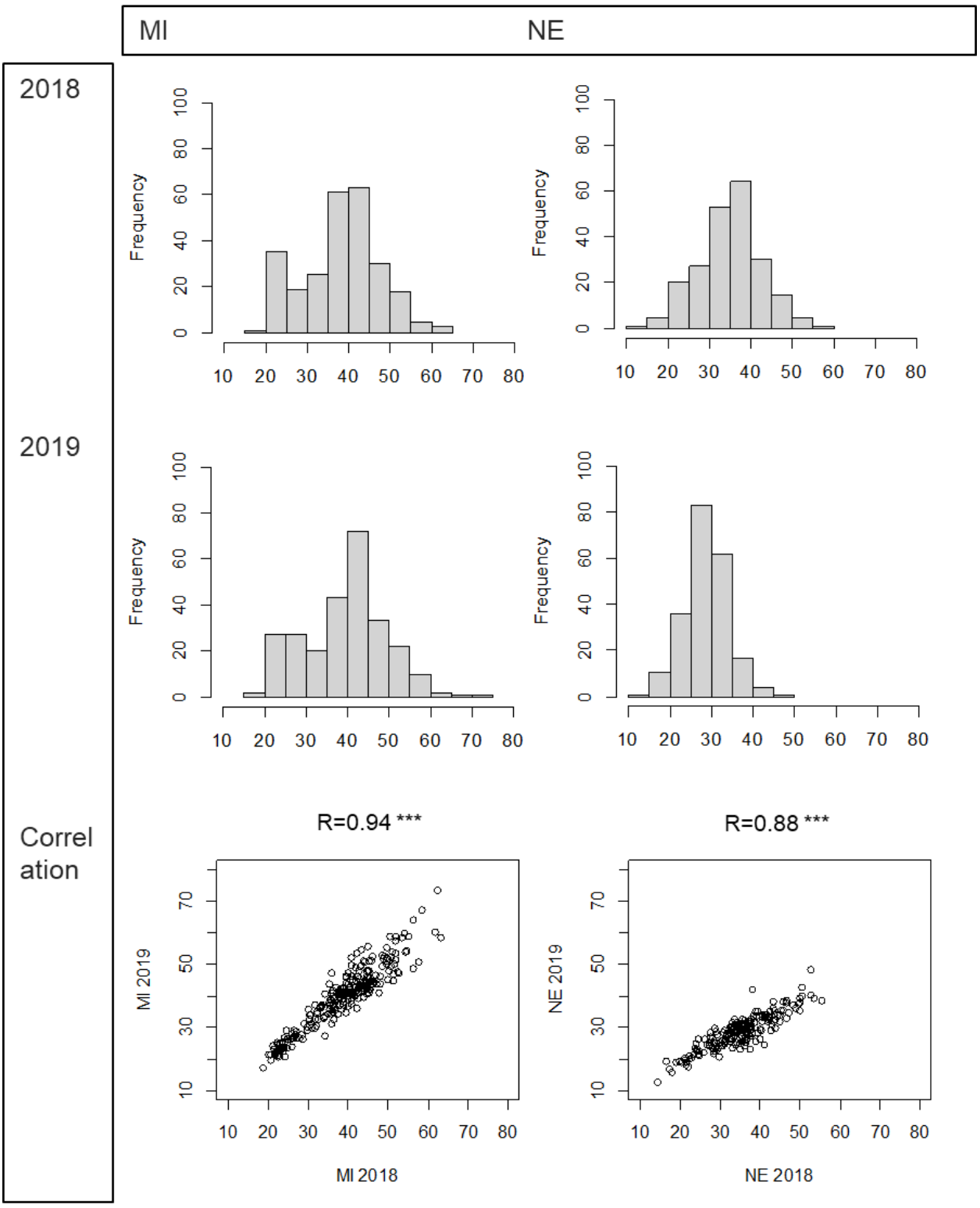


Figure S1.12 Phenotypic distribution of 100-seed weight measured in MI and NE in 2018-2019 and the correlation of the 2 years within each environment. (***: p<0.001).

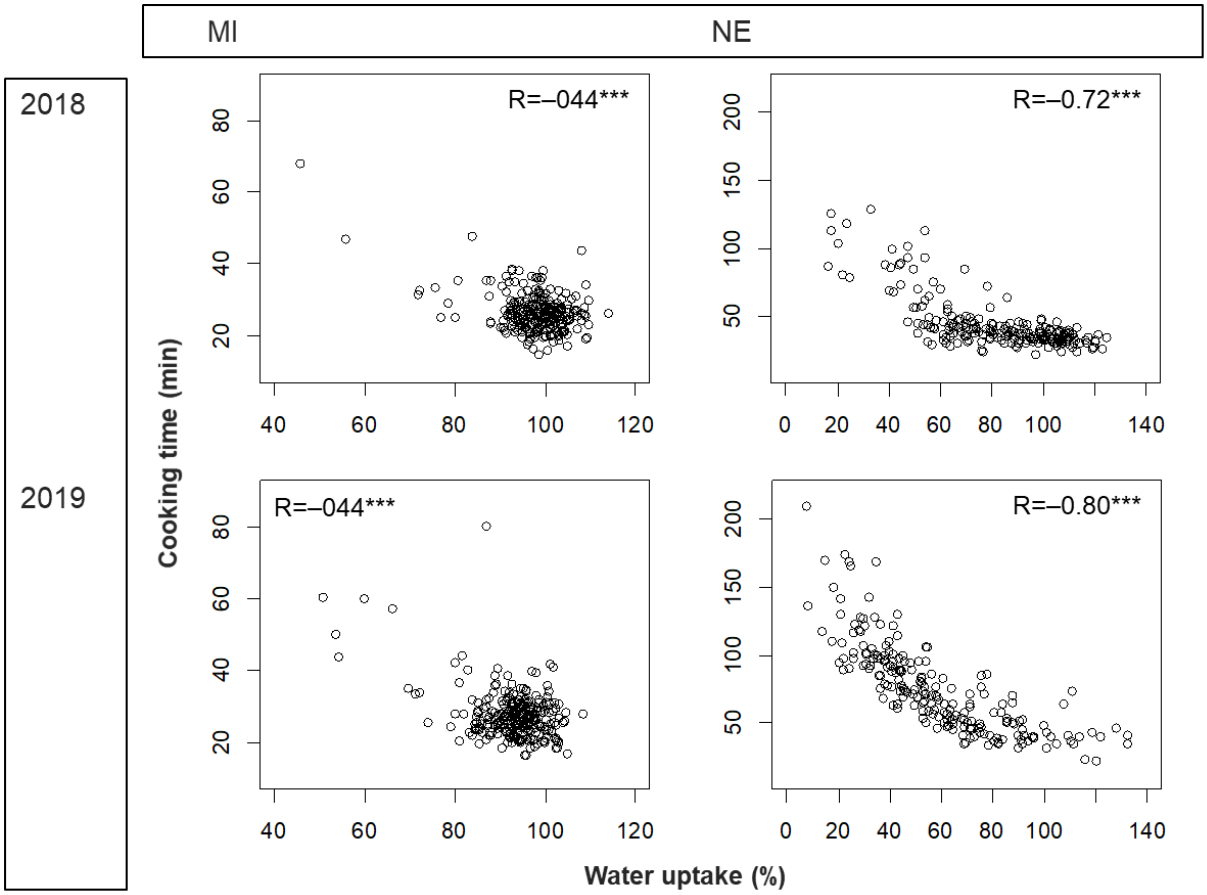


Figure S1.13 The distribution of water uptake and cooking time in MI and NE in 2018-2019. (***: p<0.001).

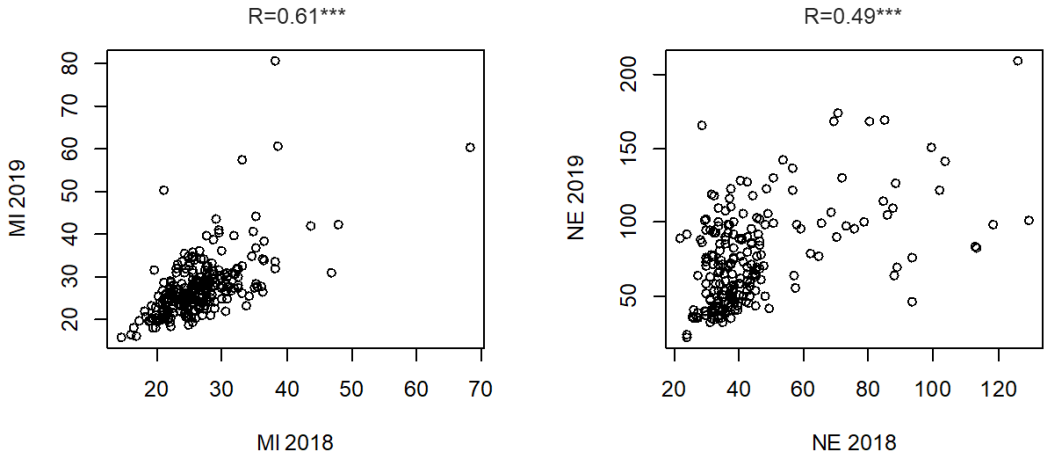


Figure S1.14 The correlation of cooking time between the years 2018 and 2019 within each environment. (***: p<0.001).

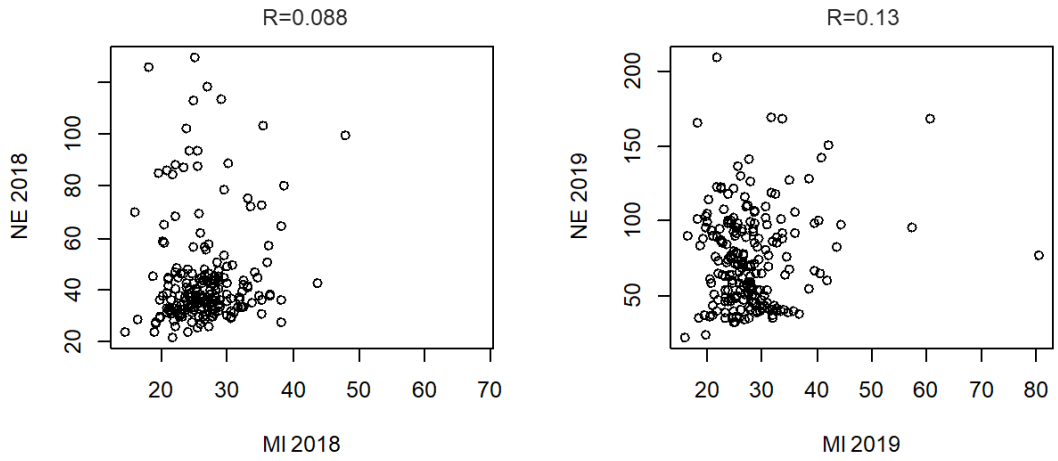


Figure S1.15 The correlation of cooking time between the two environments in 2018 and 2019.

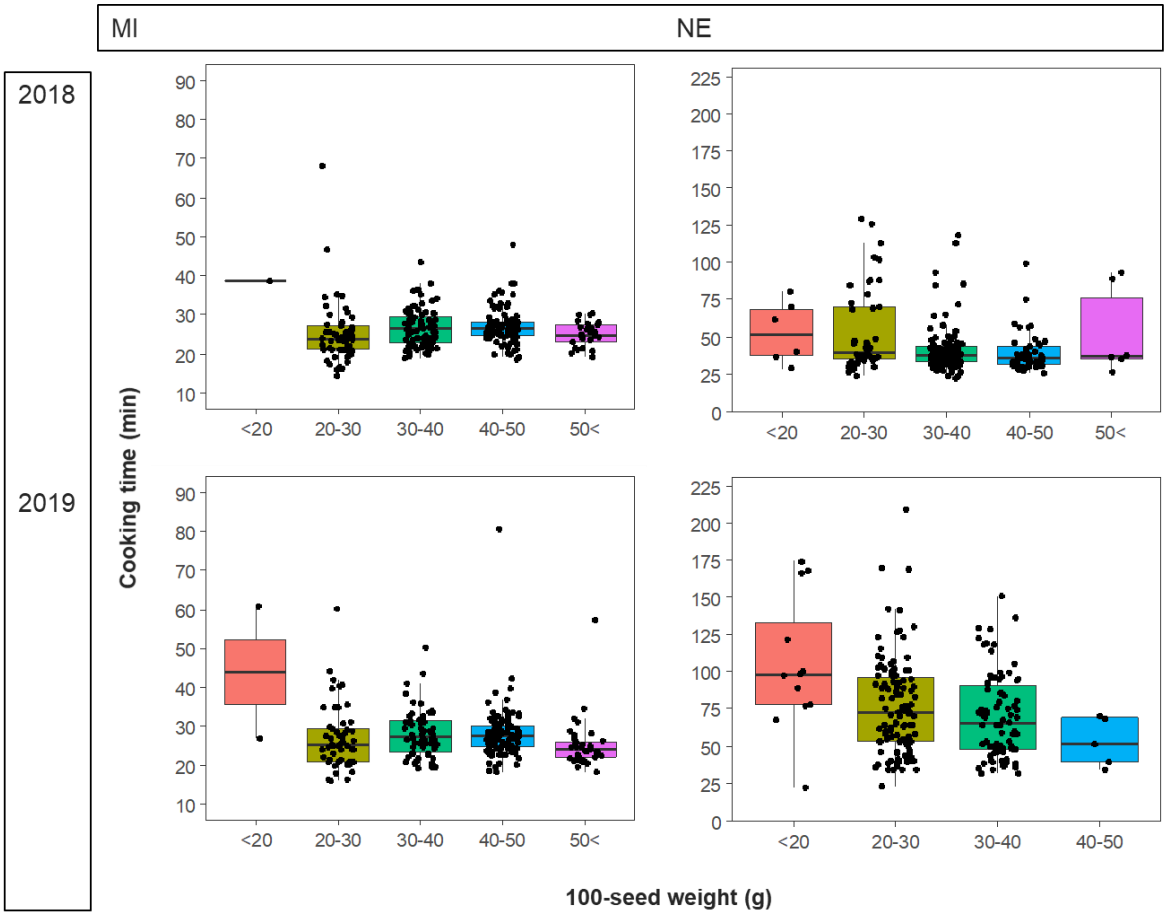


Figure S1.16 Seed size (x axis, g/100 seeds) and cooking time in MI and NE in 2018-2019.

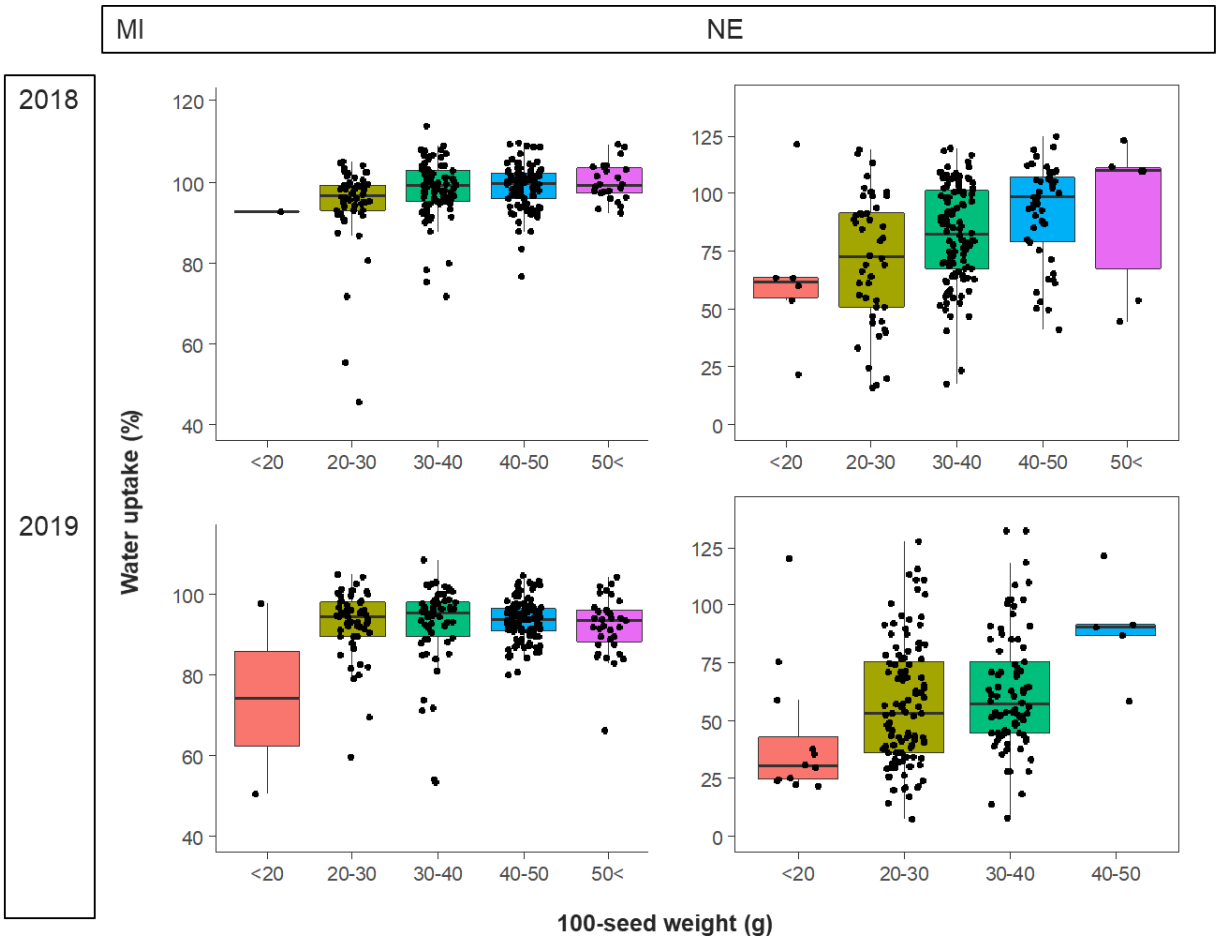


Figure S1.17 Seed size (x axis, g/100 seeds) and water uptake in MI and NE in 2018-2019.

REFERENCES

REFERENCES

Aseete, P., Katungi, E., Bonabana-Wabbi, J., Birachi, E., & Ugen, M. A. (2018). Consumer demand heterogeneity and valuation of value-added pulse products: a case of precooked beans in Uganda. *Agriculture & Food Security*, *7*(1), 51. https://doi.org/10.1186/s40066-018-0203-3

Bassett, A., Kamfwa, K., Ambachew, D., & Cichy, K. (2021). Genetic variability and genome-wide association analysis of flavor and texture in cooked beans (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*. https://doi.org/10.1007/s00122-020-03745-3

Bassett, A., Katuuramu, D. N., Song, Q., & Cichy, K. (2021). QTL mapping of seed quality traits including cooking time, flavor, and texture in a yellow dry bean (*Phaseolus vulgaris* L.) population. *Frontiers in Plant Breeding*, (in revision).

Bassett, M J. (1999). The seedcoat color genotype of "Prim" and the Manteca and Coscorron market classes of common bean. *HortScience*, *34*(2), 336–337.

Bassett, M. J. (2002). Classical and Molecular Genetic Studies of the Strong Greenish Yellow Seedcoat Color in `Wagenaar' and `Enola' Common Bean. *Journal of the American Society for Horticultural Science*, *127*(1), 50–55. http://journal.ashspublications.org/content/127/1/50.abstract

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaver, J., Zapata, M., Alameda, M., Porch, T., Rosas, J. C., Godoy-Lutz, G., & Prophete, E. (2012). Common bean improvement in the Caribbean. *2012 Global Pulse Researchers Meeting*. http://crsps.net/resource/common-bean-improvement-in-the-caribbean/

Beninger, C. W., Hosfield, G. L., & Nair, M. G. (1998). Flavonol glycosides from the seed coat of a new Manteca-type dry bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, *46*(8), 2906–2910. https://doi.org/10.1021/jf9801522

Berry, M., Izquierdo, P., Jeffery, H., Shaw, S., Nchimbi-Msolla, S., & Cichy, K. (2020). QTL analysis of cooking time and quality traits in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *133*(7), 2291–2305. https://doi.org/10.1007/s00122-020-03598-w

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) -- model food legumes. *Plant and Soil*, *252*(1), 55–128. https://doi.org/10.1023/A:1024146710611

Brouwer, I., Hartog, A. P. den, Kamwendo, M., & Heldens, M. (1996). Wood quality and wood preferences in relation to food preparation and diet composition in Central Malawi. *Ecology of Food and Nutrition*, *35*(1), 1–13. https://doi.org/10.1080/03670244.1996.9991471

Buruchara, R., Chirwa, R., Sperling, L., Mukankusi, C., Rubyogo, J. C., Mutonhi, R., & Abang, M. (2011). Development and delivery of bean varieties in Africa: the Pan-Africa Bean Research Alliance (PABRA) model. *African Crop Science Journal*, *19*(4), 227–245.

Castellanos, J. Z., Guzmán-Maldonado, H., Acosta-Gallegos, J. A., & Kelly, J. D. (1995). Effects of hardshell character on cooking time of common beans grown in the semiarid highlands of Mexico. *Journal of the Science of Food and Agriculture*, *69*(4), 437–443. https://doi.org/10.1002/jsfa.2740690406

Chigwedere, C. M., Nkonkola, C. M., Rai, S., Kyomugasho, C., Kermani, Z. J., Pallares Pallares, A., Van Loey, A. M., Grauwet, T., & Hendrickx, M. E. (2019). Cotyledon pectin molecular interconversions explain pectin solubilization during cooking of common beans (*Phaseolus vulgaris*). *Food Research International*, *116*, 462–470. https://doi.org/10.1016/j.foodres.2018.08.062

Cichy, K. A., Porch, T. G., Beaver, J. S., Cregan, P., Fourie, D., Glahn, R. P., Grusak, M. A., Kamfwa, K., Katuuramu, D. N., McClean, P., Mndolwa, E., Nchimbi-Msolla, S., Pastor-Corrales, M. A., & Miklas, P. N. (2015). A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Science*, *55*, 2149–2160. https://doi.org/10.2135/cropsci2014.09.0653

Cichy, K. A., Wiesinger, J. A., Berry, M., Nchimbi-Msolla, S., Fourie, D., Porch, T. G., Ambechew, D., & Miklas, P. N. (2019). The role of genotype and production environment in determining the cooking time of dry beans (*Phaseolus vulgaris* L.). *Legume Science*, *1*(e13). https://doi.org/10.1002/leg3.13

Cichy, K. A., Wiesinger, J. A., & Mendoza, F. A. (2015). Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *128*(8), 1555–1567. https://doi.org/10.1007/s00122-015-2531-z

Cooper, M., & DeLacy, I. H. (1994). Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics1*, *88*, 561–572.

Díaz, L. M., & Blair, M. W. (2006). Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theoretical and Applied Genetics*, *114*(1), 143–154. https://doi.org/10.1007/s00122-006-0417-9

Diaz, S., Ariza-Suarez, D., Ramdeen, R., Aparicio, J., Arunachalam, N., Hernandez, C., Diaz, H., Ruiz, H., Piepho, H.-P., & Raatz, B. (2021). Genetic architecture and genomic prediction of cooking time in common bean (*Phaseolus vulgaris* L.). *Frontiers in Plant Science*, *11:622213*. https://doi.org/10.3389/fpls.2020.622213

Dickerson, G. E. (1962). Implications of genetic-environmental interaction in animal breeding.

Animal Science, *4*(1), 47–63. https://doi.org/10.1017/S0003356100034395

Duitama, J., Quintero, J. C., Cruz, D. F., Quintero, C., Hubmann, G., Foulquié-Moreno, M. R., Verstrepen, K. J., Thevelein, J. M., & Tohme, J. (2014). An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Research*, *42*(6), e44. https://doi.org/10.1093/nar/gkt1381

Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Elia, F. M., Hosfield, G. L., Kelly, J. D., & Uebersax, M. (1997). Genetic analysis and interrelationships between traits for cooking time, water absortion, and protein and tannin content of andean dry beans. In *J. Amer. Soc. Hort. Sci.* (Vol. 122, Issue 4, pp. 512–518).

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Engleright, R., Beimiriki, M., & Hosfield, G. (1999). Determination of total dietary fiber, indigestible starch, and indigestible protein in dry bean (*Phaseolus vulgaris* L.). *Annual Report of the Bean Improvement Cooperative*, *42*, 123–124.

Erfatpour, M., Navabi, A., & Pauls, K. P. (2018). Mapping the non-darkening trait from 'Wit-rood boontje' in bean (*Phaseolus vulgaris*). *Theoretical and Applied Genetics*, *131*(6), 1331–1343. https://doi.org/10.1007/s00122-018-3081-y

Erfatpour, M., & Pauls, K. P. (2020). A R2R3-MYB gene-based marker for the non-darkening seed coat trait in pinto and cranberry beans (*Phaseolus vulgaris* L.) derived from 'Wit-rood boontje.' *Theoretical and Applied Genetics*, *133*(6), 1977–1994. https://doi.org/10.1007/s00122-020-03571-7

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fehr, W. R. (1987). *Principles of Cultivar Development*. Macmillan USA.

Felicetti, E., Song, Q., Jia, G., Cregan, P., Bett, K. E., & Miklas, P. N. (2012). Simple sequence repeats linked with slow darkening trait in pinto bean discovered by single nucleotide polymorphism assay and whole genome sequencing. *Crop Science*, *52*(4), 1600–1608. https://doi.org/10.2135/cropsci2011.12.0655

Garcia, R. A. V., Rangel, P. N., Bassinello, P. Z., Brondani, C., Melo, L. C., Sibov, S. T., & Vianello-Brondani, R. P. (2012). QTL mapping for the cooking time of common beans. *Euphytica*, *186*(3), 779–792. https://doi.org/10.1007/s10681-011-0587-7

Gepts, P., & Bliss, F. A. (1985). F1 hybrid weakness in the common bean: Differential geographic

origin suggets two gene pools in cultivated bean germplasm. *Journal of Heredity*, *76*(6), 447–450. https://doi.org/10.1093/oxfordjournals.jhered.a110142

Gepts, P., Osborn, T. C., Rashka, K., & Bliss, F. A. (1986). Phaseolin-protein variability in wild forms and landraces of the common bean(*Phaseolus vulgaris*): Evidence for multiple centers of domestication. *Economic Botany*, *40*(4), 451–468. https://doi.org/10.1007/BF02859659

Hart, J. J., Tako, E., Wiesinger, J., & Glahn, R. P. (2020). Polyphenolic profiles of yellow bean seed coats and their relationship with iron bioavailability. *Journal of Agricultural and Food Chemistry*, *68*(3), 769–778. https://doi.org/10.1021/acs.jafc.9b05663

Hirata, K., Masuda, R., Tsubokura, Y., Yasui, T., Yamada, T., Takahashi, K., Nagaya, T., Sayama, T., Ishimoto, M., & Hajika, M. (2014). Identification of quantitative trait loci associated with boiled seed hardness in soybean. *Breeding Science*, 64(4), 362–370. https://doi.org/10.1270/jsbbs.64.362

Hooper, S., Wiesinger, J. A., Echeverria, D., Thompson, H. J., Brick, M. A., Nchimbi-Msolla, S., & Cichy, K. A. (2016). Carbohydrate profile of a dry bean (*Phaseolus vulgaris* L.) panel encompassing broad genetic variability for cooking time. *Cereal Chemistry Journal*, *94*(1), 135–141. https://doi.org/10.1094/CCHEM-04-16-0126-FI

Hosfield, G., Bennink, M., Beninger, C., Engleright, R., & Ospina, M. (1998). Variability for starch digestibility in dry bean (*Phaseolus vulgaris*). *HortScience*, *33*, 472. https://journals.ashs.org/hortsci/view/journals/hortsci/33/3/article-p443.xml

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, *8*(2). https://doi.org/10.1093/gigascience/giy154

Jacinto-hernandez, C., Azpiroz-rivero, S., Acosta-gallegos, J. A., Hernandez-sanchez, H., & Bernal-lugo, I. (2003). Genetic analysis and random amplified polymorphic DNA markers associated with cooking time in common bean. *Crop Science*, *43*, 329–332.

Katuuramu, D. N., Luyima, G. B., Nkalubo, S. T., Wiesinger, J. A., Kelly, J. D., & Cichy, K. A. (2020). On-farm multi-location evaluation of genotype by environment interactions for seed yield and cooking time in common bean. *Scientific Reports*, *10*(1), 3628. https://doi.org/10.1038/s41598-020-60087-2

Kelly, J. D., Awale, H. E., Wiersma, A. T., Cichy, K. A., & Wright, E. M. (2021). Registration of 'Yellowstone' yellow bean. *Journal of Plant Registrations*, 1–6. https://doi.org/10.1002/plr2.20075

Kelly, J. D., Wright, E. M., & Wiersma, A. (2018). 2018 Dry bean yield trials. In Michigan State University AgBioResearch (Ed.), *2018 Research Report Saginaw Valley Research & Extension* (pp. 15–51).

Kelly, J. D., Wright, E. M., & Wiersma, A. (2019). 2019 Dry bean yield trials. In Michigan State University AgBioResearch (Ed.), *2019 Research Report Saginaw Valley Research &*

Extension (pp. 16–52).

Kilimo Trust. (2012). Development of Inclusive Markets in Agriculture and Trade (DIMAT). *Undp*, 1–48. http://www.undp.org/content/dam/uganda/docs/UNDPUg_ PovRed_Value Chain Analysis Report Honey 2013 Report.pdf

Koenig, R., & Gepts, P. (1989). Allozyme diversity in wild *Phaseolus vulgaris*: Further evidence for two major centers of genetic diversity. *Theoretical and Applied Genetics*, *78*(6), 809–817. https://doi.org/10.1007/BF00266663

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Leakey, C. L. A. (1992). Breeding on the C and J and B loci for modification of bean seedcoat flavonoids with the objective of improving food acceptability. *Annual Report of the Bean Improvement Cooperative*, *35*, xiii–xvii.

Lee, T.-H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, *15*(1), 162. https://doi.org/10.1186/1471-2164-15-162

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.5-1*. https://cran.r-project.org/package=emmeans

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Lobaton, J. D., Miller, T., Gil, J., Ariza, D., de la Hoz, J. F., Soler, A., Beebe, S., Duitama, J., Gepts, P., & Raatz, B. (2018). Resequencing of Common Bean Identifies Regions of Inter–Gene Pool Introgression and Provides Comprehensive Resources for Molecular Breeding. *The Plant Genome*, *11*(2). https://doi.org/10.3835/plantgenome2017.08.0068

Mainviel, R. (2019). *Agronomic performance and genetic diversity of common bean (*Phaseolus vulgaris*) varieties in Haiti* [The University of Florida]. https://ufdc.ufl.edu/UFE0055877/00001

Martín-Cabrejas, M. A., Esteban, R. M., Perez, P., Maina, G., & Waldron, K. W. (1997). Changes in physicochemical properties of dry beans (*Phaseolus vulgaris* L.) during long-term storage. *Journal of Agricultural and Food Chemistry*, *45*(8), 3223–3227. https://doi.org/10.1021/jf970069z

Martínez-Manrique, E., Jacinto-Hernández, C., Garza-García, R., Campos, A., Moreno, E., & Bernal-Lugo, I. (2011). Enzymatic changes in pectic polysaccharides related to the beneficial effect of soaking on bean cooking time. *Journal of the Science of Food and Agriculture*, *91*(13), 2394–2398. https://doi.org/10.1002/jsfa.4474

McClean, P. E., Mamidi, S., McConnell, M., Chikara, S., & Lee, R. (2010). Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics*, *11*(1), 184. https://doi.org/10.1186/1471-2164-11-184

Mishili, F. J., Temu, A., Fulton, J., & Lowenberg-DeBoer, J. (2011). Consumer preferences as drivers of the common bean trade in Tanzania: A marketing perspective. *Journal of International Food & Agribusiness Marketing*, *23*(2), 110–127. https://doi.org/10.1080/08974438.2011.558761

Molnar, S. J., Charette, M., & Cober, E. R. (2012). Mapping quantitative trait loci for water uptake in a recombinant inbred line population of natto soybean. *Canadian Journal of Plant Science*, *92*(2), 257–266. https://doi.org/10.4141/cjps2011-122

Oladzad, A., Porch, T., Rosas, J. C., Moghaddam, S. M., Beaver, J., Beebe, S. E., Burridge, J., Jochua, C. N., Miguel, M. A., Miklas, P. N., Ratz, B., White, J. W., Lynch, J., & McClean, P. E. (2019). Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. *G3: Genes|Genomes|Genetics*, *9*(6), 1881–1892. https://doi.org/10.1534/g3.119.400072

Ott, A., Trautschold, B., & Sandhu, D. (2011). Using microsatellites to understand the physical distribution of recombination on soybean chromosomes. *PLoS ONE*, *6*(7), e22306. https://doi.org/10.1371/journal.pone.0022306

Pallottini, L., Garcia, E., Kami, J., Barcaccia, G., & Gepts, P. (2004). The genetic anatomy of a patented yellow bean. *Crop Science*, *44*(3), 968–977. https://doi.org/10.2135/cropsci2004.9680

Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., & Ferreira, J. J. (2010). Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *120*(7), 1367–1380. https://doi.org/10.1007/s00122-010-1261-5

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959. http://www.genetics.org/content/155/2/945.abstract

Pritchard, J. K., Wen, X., & Falush, D. (2009). *Documentation for structure software: Version 2.3* (pp. 1–39). https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf

Proctor, L. (1999). *Field bean culticar called Enola* (Patent No. US5,894,079).

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Rainey, K. M., & Griffiths, P. D. (2005). Differential response of common bean genotypes to high temperature. *Journal of the American Society for Horticultural Science*, *130*(1), 18–23. https://doi.org/10.21273/JASHS.130.1.18

Ridley, B. L., O'Neill, M. A., & Mohnen, D. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, *57*(6), 929–967. https://doi.org/10.1016/S0031-9422(01)00113-3

Rockland, L. B., & Jones, F. T. (1974). Scanning electron microscope studies on dry beans. Effects of cooking on the cellular structure of cotyledons in rehydrated large lima beans. *Journal of Food Science*, *39*(2), 342–346. https://doi.org/10.1111/j.1365-2621.1974.tb02890.x

Sandhu, K. S., You, F. M., Conner, R. L., Balasubramanian, P. M., & Hou, A. (2018). Genetic analysis and QTL mapping of the seed hardness trait in a black common bean (*Phaseolus vulgaris*) recombinant inbred line (RIL) population. *Molecular Breeding*, *38*(3), 34. https://doi.org/10.1007/s11032-018-0789-y

Sichilima, T., Mapemba, L., & Tembo, G. (2016). Drivers of dry common beans trade in Lusaka, Zambia: A trader's perspective. *Sustainable Agriculture Research*, *5*(2), 15–26. https://doi.org/10.5539/sar.v5n2p15

Smith, L. P., Ng, S. W., & Popkin, B. M. (2013). Trends in US home food preparation and consumption: analysis of national nutrition surveys and time use studies from 1965–1966 to 2007–2008. *Nutrition Journal*, *12*(1), 45. https://doi.org/10.1186/1475-2891-12-45

Soltani, A., Walter, K. A., Wiersma, A. T., Santiago, J. P., Quiqley, M., Chitwood, D., Porch, T. G., Miklas, P., McClean, P. E., Osorno, J. M., & Lowry, D. B. (2021). The genetics and physiology of seed dormancy, a crucial trait in common bean domestication. *BMC Plant Biology*, *21*(1), 58. https://doi.org/10.1186/s12870-021-02837-6

Sones, D. (2015). Soya Njano is the bean for home consumption. *Our Blog: The Inside Story; Africa Soil Health Consortium*. https://africasoilhealth.cabi.org/2015/09/29/soya-njano-is-the-bean-for-home-consumption/

Stanley, D. W. (1992). A possible role for condensed tannins in bean hardening. *Food Research International*, *25*(3), 187–192. https://doi.org/10.1016/0963-9969(92)90136-S

Tumeo, M., Mapemba, L., Edriss, A. K., & Phiri, H. (2017). *Consumer choice of dry common beans in Malawi: The case of Lilongwe City* (No. 19; MaSSP Working Paper). http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/131395

US Department of Agriculture Agricultural Marketing Service. (2019). "Patron" yellow bean. *Plant Variety Protection Certificate*, 1–27. https://apps.ams.usda.gov/cms/adobeimages/201700233.pdf

Venables, W., & Ripley, B. (2002). *Modern applied statistics with S* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Voysest, O. (2012). Yellow beans in Latin America. *Annual Report of the Bean Improvement Cooperative*, xii–xviii. http://bic.css.msu.edu/_pdf/Reports/BIC_2012_Annual_Report.pdf

Wang, N., & Daun, J. K. (2005). Determination of cooking times of pulses using an automated Mattson cooker apparatus. *Journal of the Science of Food and Agriculture*, *85*(10), 1631–1635. https://doi.org/10.1002/jsfa.2134

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wiesinger, J. A., Cichy, K. A., Glahn, R. P., Grusak, M. A., Brick, M. A., Thompson, H. J., & Tako, E. (2016). Demonstrating a nutritional advantage to the fast-cooking dry bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, *64*(45), 8592–8603. https://doi.org/10.1021/acs.jafc.6b03100

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

Wortmann, C. S., Kirkby, R. A., Eledu, C. A., & Allen, D. J. (1998). Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa. *CIAT Publication*, *297*, 131. http://hdl.handle.net/10568/54312

Yin, L. (2020). *CMplot: Circle Manhattan Plot. R package version 3.6.2*. https://cran.r-project.org/package=CMplot

CHAPTER 2:

SEED COAT COLOR GENETICS AND G×E IN A YELLOW BEAN COLLECTION VIA IMAGE ANALYSIS PAIRED WITH MACHINE-LEARNING AND GWAS

[Submitted for publication in Plant Genome]

Seed Coat Color Genetics and G×E in a Yellow Bean Collection via Image Analysis Paired with Machine-Learning and GWAS

Rie Sadohara[1], Yunfei Long[2], Paulo Izquierdo[1], Carlos A. Urrea[3], Daniel Morris[2], Karen Cichy[1,4]

1. Department of Plant, Soil and Microbial Sciences, Michigan State University, 1066 Bogue St., East Lansing, MI 48824, USA
2. Department of Electrical and Computer Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, MI 48824, USA
3. Panhandle Research & Extension Center, University of Nebraska, 4502 Ave I, Scottsbluff, NE 69361, USA
4. Sugarbeet and Bean Research Unit, USDA-ARS, 1066 Bogue St., East Lansing, MI 48824, USA

Abstract

Common bean (*Phaseolus vulgaris* L.) is consumed worldwide, and consumers show solid regional preferences for specific seed appearance. Colors of the seed coat, hilum ring, and corona are all important, along with susceptibility to postharvest darkening, which decreases seed value. This study aimed to characterize a collection of 295 yellow bean genotypes for seed appearance and postharvest darkening behavior, evaluate genotype × environment effects and map those traits via genome-wide association analysis. The yellow beans were grown for two years in Michigan and Nebraska, USA, and were evaluated for L*a*b*, postharvest darkening, and hilum ring and corona colors. A model to exclude the hilum ring and corona of the seeds, black background, and light reflection was developed by using machine learning, allowing for targeted and efficient L*a*b* value extraction from the seed coat. The genotype × environment effects were significant for the color values, and MI-grown seeds had darker seeds than NE-grown seeds. SNPs were

associated with L* and hilum ring color on Pv10 near the *J* gene involved in mature seed coat color and hilum ring color. A SNP on Pv07 associated with L*, a*, postharvest darkening, and hilum ring and corona colors was near the *P*, the ground factor gene for seed coat color expression. The machine learning-aided model used to extract color values from the seed coat, the wide variability in seed morphology traits, and the associated SNPs will provide tools for future breeding and research efforts to meet consumers' expectations for bean seed appearance.

Introduction

Common bean (*Phaseolus vulgaris* L.) is a nutrient-dense food consumed worldwide, with an annual world production of 29 million tons (average of 2016-2019; FAOSTAT, 2021). Beans are an excellent source of protein, dietary fiber, and micronutrients and offer numerous health benefits, including prevention and/or control of cardiometabolic diseases and certain types of cancer (Hayat et al., 2014; Messina, 2014). The seed appearance of beans is diverse with various shapes, sizes, colors, and patterns, the combinations of which form specific bean market classes (seed types) (González et al., 2006; Uebersax & Siddiq, 2012; Wortmann et al., 1998). Therefore, bean breeders develop varieties with morphological characteristics that meet the market demands in their target regions (Beebe, 2020; Castellanos et al., 1996; O. Voysest et al., 1994). In addition to the primary seed color and pattern, the hilum ring and corona color are also essential parts of seed appearance.

Since early in the 20th century, extensive research has been conducted to identify genes responsible for producing bean seed coat color. Prakken (1972, 1974) consolidated initial findings and developed a model for color expression in the bean seed coat, which is primarily accepted to this date. The model groups color-related genes into three categories: a ground factor gene, color genes, and color-modifying genes. A ground factor gene, *P*, is necessary for any seed color

expression, and genotypes with homozygous recessive *p* produce a white seed coat. *P* gene was later found to encode a transcription factor required for initiating flavonoid biosynthesis (McClean et al., 2018). Seed colors include *C, D,* and *J* genes. *C* gene gives a pale greenish-yellow seed coat and mottling patterns. *C* is tightly linked with other genes forming 'complex [*C R Prp*] locus', which are involved in different color and pattern expressions. (M. J. Bassett, 2007). *D* gene gave a brown hilum ring and was later found to be an allele of *Z* which is also involved in the partly colored seed coat (M. J. Bassett et al., 1999). *J* gene is involved in multiple traits such as seed coat color formation during maturation, brown hilum ring, seed coat shininess, and postharvest darkening (M. J. Bassett, 1996; Prakken, 1970). Postharvest darkening refers to a phenomenon where seed color of some market classes darkens over time during storage. Postharvest darkening is an important trait for bean producers because consumers consider darkened seeds as old and long-cooking, so darkened seeds may be discounted (Junk-Knievel et al., 2008). Previous studies have shown that two unlinked interacting genes control postharvest darkening: *J* and *sd,* with the *J_* genotype presenting postharvest darkening and *jj* genotype non-darkening, and recessive *sdsd* paired with *J_* showing slow darkening behavior (Elsadr et al., 2011; Junk-Knievel et al., 2008). *sd* has been confirmed to be an allele of the ground factor *P* and is denoted as p^{sd} (N. S. Islam et al., 2020). Color-modifying genes do not confer a color by themselves but intensify color expression of the color genes. Color-modifying genes include *G* for yellow-brown, *B* for greenish yellow-brown, *V* for purple to black, *Rk* for recessive red, and *Gy* for greenish-yellow (M. J. Bassett, 2002; M. J. Bassett et al., 2010; Beninger & Hosfield, 1999). The mechanism of color expression is complicated; some of the genes are multi-allelic, and they interact with each other to collectively determine the colors and patterns of seed coat, hilum ring, corona, pods, and flowers (M. J. Bassett, 2007).

Although seed coat colors could be treated as a qualitative trait controlled by major genes, more recent research has quantitatively measured colors using CIE L*a*b* values (CIE International Commission on Illumination, 2004). CIE L*a*b* expresses colors with three values, L*, a*, and b*, which measure lightness, greenness-redness, and blueness-yellowness, respectively defined as axes in a three-dimensional color space. L*a*b* values are perceptually uniform, meaning that the degree of difference between two colors corresponds to the Euclidean distance of them in the L*a*b* space (León et al., 2006). Expressing colors using numerical values enabled an objective measurement of color rather than a subjective description. Traditionally, colorimeters were used to measure color values, but image-based computer-vision technology has been used (Wu & Sun, 2013). Image-based color measurement provides a better representation of samples because it can obtain L*, a*, and b* values from each pixel of an image, and a larger food surface area can be sampled compared to conventional colorimeters (León et al., 2006). Depending on the type of samples, images can contain pixels that do not represent the specimen. In the case of bean seed color, bean images would contain white hilum and a hilum ring and corona, which surround the hilum, the background, and potentially light reflection because of the spherical nature of the seeds. Measuring color values without excluding those could result in differences in color values due to darker hilum ring and/or corona despite the same seed coat color of two bean samples. Machine-learning has been used to classify target features of plant images (Amatya et al., 2016; Ferentinos, 2018; Grinblat et al., 2016), presenting potentials of applying this technology to segmenting seed coat from non-targets in bean seed images.

Numerous polyphenolic compounds exist in the seed coat of common beans, such as flavonoids and proanthocyanidins (condensed tannins) (Yang et al., 2018). Among them, flavonoids are considered to impart bean seed color (Beninger et al., 1998; Feenstra, 1960).

Flavonoids present in the bean seed coat can be sub-classified into flavonols, flavanols, and anthocyanidins (Reddy et al., 1985). Anthocyanidins are glycosides of anthocyanins, including cyanidin, pelargonidin, and delphinidin, and are found in pink, red, and black beans (Rodríguez Madrera et al., 2020; Takeoka et al., 1997). Proanthocyanidins are dimers or polymers of flavanols, including catechin and epicatechin (Dixon et al., 2005). Proanthocyanidins produce brown deposits upon oxidation in pinto beans and are associated with postharvest darkening (Beninger & Hosfield, 2003; Marles et al., 2008). Proanthocyanidins are also involved in plant defense against pests and pathogens (F. M. A. Islam et al., 2003; Winkel-Shirley, 2001). In addition to seed coat color and plant defense responses, the concentrations of polyphenolic compounds influence quality attributes of beans such as cooking time, digestibility, and iron bioavailability (Bressani & Elías, 1979; Elia et al., 1997; Petry et al., 2010; Tako et al., 2014).

Yellow beans are an important market class in East and South Africa and Latin America (Voysest, 2012; Wortmann et al., 1998). Yellow beans are diverse in seed morphology (**Figure 2.1**), and at least a dozen yellow-seeded market classes exist in Latin America alone (Voysest, 2012). One of the valuable attributes of some yellow beans is that they do not exhibit postharvest darkening. The susceptibility to postharvest darkening of yellow beans is mainly unknown except that genetic variability and a QTL for postharvest darkening near the *J* gene were reported among Manteca and Green-yellow types and their progeny population (A. Bassett et al., 2021). Postharvest darkening is known to occur in carioca beans, which have light brown stripes on a cream background similar to pinto beans (Junk-Knievel et al., 2008); therefore, some yellow beans may be susceptible to postharvest darkening. Although high heritability estimates (73-99%) were reported for color values with populations derived from yellow-seeded beans (Arns et al., 2018; A. Bassett et al., 2021; Possobom et al., 2015; Ribeiro et al., 2019), growing conditions could

influence their seed coat colors (Kelly et al., 2021; Possobom et al., 2015). Thus, postharvest darkening behavior and the G×E effects on the seed color of the YBC should be evaluated to develop bean varieties that meet the consumers' expectations in various bean production regions.

Evidence is accumulating that a lemon-yellow seeded type, Manteca, is a promising breeding material because of its high digestibility, short cooking time, and high iron bioavailability (Cichy et al., 2015; Engleright et al., 1999; Hart et al., 2020; Hooper et al., 2016; J. Wiesinger et al., 2018; J. A. Wiesinger et al., 2019). Other yellow beans' convenience and nutritional qualities have yet to be assessed despite their wide morphological diversity. Therefore, the Yellow Bean Collection (YBC) has been assembled with yellow-seeded *P. vulgaris* germplasm collected globally (Sadohara et al., 2021); **Figure S2.6**). Quantitative characterization of the seed coat colors of the YBC would enable an objective evaluation of their appearance, an important characteristic to consumers.

The objectives of this study were to (1) assess the genetic diversity of the Yellow Bean Collection of 295 yellow bean genotypes for seed coat color values via a machine-learning aided procedure; (2) to assess the postharvest darkening behavior of the YBC; and (3) investigate the genetic control and the G×E on these traits to aid breeding of beans with desired morphological, nutritional, and culinary characteristics.

Materials and Methods

Plant materials

A yellow bean diversity panel, the Yellow Bean Collection (YBC), comprising 295 genotypes was used in this study. The YBC was grown in two environments, MI and NE, USA in 2018 and 2019. Detailed experimental design is described by Sadohara et al. (2021). The numbers of genotypes

phenotyped for color, postharvest darkening, and hilum ring and corona colors are shown in **Table S2.6**.

CIE L*a*b* of bean seed coat

The YBC contained solid-colored and patterned seed types, but the solid-colored types were used for seed coat color measurements. Namely, the genotypes used for color measurements were: 37 Amarillo dark, 44 Amarillo light, 6 beige, 7 brown, 83 Canary, 6 green-yellow including Njano, 11 Manteca, 56 Mayocoba, and 3 white genotypes. Solid-colored one layer of intact whole seeds were placed in a 70 mm × 70 mm box with matte black walls. Images of bean seeds from each of the two field replications were taken by using an image acquisition system described by Mendoza et al. (2017) with the shutter speed set to 1/100 s. To extract bean color from captured images, a binary pixel classification was defined in which pixels representative of bean color are one category, and the remaining pixels, i.e., background, hilum ring, corona, and specular reflection areas, are the second category. The pyramid convolutional neural network developed in Long et al. (2019) to detect bean seed coat splits, was retrained for this color pixel segmentation task. To train the network, semantic labels for 24 images of various bean types and colors within the YBC were created manually and augmented by rotation and flipping. The network achieved 0.95 average precision in bean pixel classification on a test set and so was able to effectively eliminate non-desired pixels from images, as illustrated in **Figure 2.2**.

Bean images were downsampled to 432 × 432 pixels using a custom MATLAB script (MATLAB, 2020), and the segmentation model was applied. CIE L*, a*, and b* values were extracted from valid pixels in the images by scikit-image in Python (van der Walt et al., 2014). The L*a*b* values were calibrated using an image of Macbeth Color Checker Chart (X-Rite,

Grand Rapids, MI, USA) taken under the same conditions as the YBC at the start of image acquisition each year. The 24 color patches in the Macbeth Color Checker Chart image were cut out into 100 × 100-pixel squares using ImageJ 1.52v (Schneider et al., 2012), and the L*a*b* values were extracted from each patch by the scikit-image in Python. The extracted L*a*b* values and the standard values of the patches were used to build a partial least squares (PLS) model by the PLS package (Mevik et al., 2020) in R (R Core Team, 2017). Specifically, SIMPLS, a faster and improved PLS model (de Jong, 1993) with three components, was used to calibrate L*a*b* values of the YBC bean images, and calibration was carried out each year to accommodate potential changes in lighting intensity over the one year.

To evaluate the difference in seed coat colors between the environments, ΔL*, Δa*, and Δb* were calculated as the mean L*a*b* values in MI subtracted by those in NE for each genotype. ΔE was computed as $\Delta E = \sqrt[2]{\Delta \mathrm{L}*^2 + \Delta \mathrm{a}*^2 + \Delta b*^2}$.

Postharvest darkening

Postharvest darkening of MI 2019 seeds was evaluated via a UV test according to the method developed by Junk-Knievel et al. (2007) as follows: approximately 10 seeds per sample per field replication were placed in a clear plastic seed tray. The tray held 60 samples in total. Trays containing seeds were placed approximately 10 cm under a UV light for 96 h. After that time period, the seeds were scored for darkening as 0 for non or slow darkening and 1 for darkening. Some of the dark-seeded genotypes such as brown and Amarillo (dark) types were excluded from this evaluation.

Hilum ring and corona colors

The hilum ring and corona colors were evaluated of the MI 2019 seeds. The hilum ring colors were classified into two categories: 0 for yellow and 1 for dark. The corona colors were classified into two categories: 0 for light/yellow, and 1 for dark.

Statistical analyses

The L*a*b* values were transformed using the Box-Cox method (Box & Cox, 1964) by using the MASS package (Venables & Ripley, 2002) in R. The genotype, environment, and year effects on the color values were estimated by the following mixed linear model: $Y_{ijkl} = \mu + G_i + E_j + Y_k + EY_{jk} + GE_{ij} + GY_{ik} + GEY_{ijk} + B(EY)_{jkl} + \varepsilon_{ijkl}$, where: Y_{ijkl} is the phenotypic value of the *i*th YBC genotype grown in the *l*th block of the *j*th environment in the *k*th year, μ is the grand mean, G_i is a random effect of the *i*th genotype, and E_j and Y_k are fixed effects of the *j*th location and the *k*th year, respectively. EY_{jk}, GE_{ij}, and GY_{ik} are two-way interaction terms, GEY_{ijk} is a three-way interaction term, $B(EY)_{jkl}$ is a random effect of the *l*th block nested in the *j*th environment and the *k*th year, and ε_{ijkl} is the error term. The variance components of the model were used to estimate the broad-sense heritability of the phenotypic values (Fehr, 1987). To minimize the environment and year effect, best linear unbiased estimate (BLUE) was calculated for the L*, a*, and b* values by setting the genotype as a fixed effect and all other terms as random effects using the emmeans package (Lenth, 2021) in R. All the models were fitted using the lme4 (Bates et al., 2015) and the lmerTest (Kuznetsova et al., 2017) packages in R. The BLUE values were used for principal component analysis (PCA) and genome-wide association analysis. PCA was performed using prcomp function and was visualized using the factoextra package (Kassambara & Mundt, 2020) in R.

Genome-wide association analysis

Genome-wide association analyses were carried out using the BLUE of the CIE L*a*b*, coded darkening phenotype (0: non-darkening,1: darkening), hilum ring color (0: yellow, 1: dark), and corona color (0: light/yellow, 1: dark). In total, 253 genotypes for L*a*b*, 206 genotypes for postharvest darkening, and 258 genotypes for hilum ring and corona colors were used. The 295 genotypes of the YBC were sequenced via Genotyping-By-Sequencing technology (Elshire et al., 2011), generating a base SNP set of 417,142 SNPs by NGSEP (Duitama et al., 2014). Details of the genotyping pipeline are described by Sadohara et al. (2021). The base SNP set was filtered to include the 253, 206, and 258 phenotyped individuals. The SNPs were selected if they were biallelic, had no heterozygous calls, had a genotyping quality score > 40, were not on scaffolds, were outside the repetitive regions of the common bean genome (ver. 2; Lobaton et al., 2018), had < 20% missing data, and had > 5% minor allele frequency (MAF). As a result, 2,277 SNPs for L*a*b*, 2,278 SNPs for postharvest darkening, and 2,315 SNPs for hilum ring and corona colors were retained and used for genome-wide association analyses with the GAPIT package (Lipka et al., 2012) in R. Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) method was used to detect associations between the phenotypes and the genotypes (Huang et al., 2019). A false discovery rate (FDR) at α=0.05 was used to call associations significant. QQ plots and Manhattan plots were generated by using the CMplot package (Yin, 2020) in R.

Results and discussion

Phenotypic diversity of seed coat, hilum ring, and corona colors and postharvest darkening behavior

Employing the machine-learning technology, seed coat was successfully separated from hilum and corona of seeds, black background, and light reflection after feeding 24 hand-labeled images for training. The model correctly distinguished pixels of seed coat of more than 1,800 images of the YBC genotypes without manual operation. The automation allows for more accurate measurement of seed coat color and saves a tremendous amount of time and labor that would have been spent on manual processing of images. Automated separation of seed coat and hilum ring and corona areas are substantial because they are all under different genetic controls (M. J. Bassett, 2007). Bean seed colors have been measured traditionally with colorimeters (Arns et al., 2018; Erfatpour et al., 2018; Possobom et al., 2015) and more recently with ImageJ (A. Bassett et al., 2021; Bornowski et al., 2020; Varga et al., 2019), but the seed coat, hilum ring, and corona colors were not separated. Thus, this study has provided a novel, automated method to obtain color values from seed coat pixels.

The YBC consists of 295 yellow-colored beans with solid-colored or patterned seed coats. The L*a*b* color values were evaluated for 252 genotypes in MI in 2018, 250 in MI 2019, 209 in NE 2018, and 195 in NE, depending on the seed availability and adaptation (**Table S2.6**). A wide diversity in the seed coat color values were observed among the YBC genotypes: L* values ranged from 27.6 to 86.0, a* values from −5.9 to 35.0, and b* values from 11.9 to 63.6 (**Table 2.1**). The two-year average L*a*b* values were highly correlated between the environments, but the correlation coefficient was slightly lower for the b* values (R = 0.88, $p < 0.001$), which measures yellowness, than L* and a* (R = 0.98, $p < 0.001$ for both; **Figure S2.7**). L* and a* values were

negatively correlated (R = −0.69 in MI and R = −0.73 in NE), and L* and b* values were moderately positively correlated (R = 0.5 in MI, R = 0.3 in NE).

A total of 206 genotypes excluding dark-colored seeds were evaluated for susceptibility to postharvest darkening by the UV test. As a result, 71 were non or slow darkening, and 135 genotypes were darkening (**Table 2.2**). The majority of Amarillo light and canary genotypes were darkening. All the Manteca were non or slow darkening. 'Prim' Manteca bean was reported to contain no tannins (Beninger et al., 1998) and was *jj* genotype at the *J* locus, which would result in non-darkening (M. J. Bassett, 1999); therefore, Mantecas are expected to be non-darkening. For Mayocoba, 39 out of the 54 genotypes were slow or non-darkening, and 14 were U.S. breeding lines. Because consumers highly value non and slow darkening seeds as it is associated with freshness and shorter cooking time, another important trait (Wiesinger et al., 2021; Erfatpour et al., 2018; Silva et al., 2018), information on the postharvest darkening behavior of the YBC would be useful in developing and selecting yellow beans that meet consumers' expectations.

A total of 258 YBC genotypes grown in MI 2019 were classified based on their hilum ring and corona colors. **Tables 2.3 and 2.4** show the number of genotypes in each category of hilum ring and corona colors by seed type. All the Manteca, Mayocoba, and white beans had a yellow hilum ring, and Amarillos, beige, brown, Canary, and Green-yellow types had a dark hilum ring. For corona colors, Canary and Manteca had dark corona, and Mayocoba had light/yellow corona (**Table 2.4**), consistent with their characteristic seed appearance. Some Amarillos had light/yellow corona, and others had dark corona. Fifty-three out of the 71 non or slow-darkening genotypes had a yellow hilum ring, and the other 18 had a dark hilum ring (**Table S2.7**). *J* gene is involved in both postharvest darkening and dark hilum ring color, and individuals with *jj* genotype are non-darkening (Elsadr et al., 2011; Junk-Knievel et al., 2008). The *Z* gene is also involved in hilum

ring color expression, and *Z* produces a dark hilum ring regardless of the *J* genotype (M. J. Bassett et al., 1999). Therefore, the slow or non-darkening 21 genotypes that had a dark hilum ring must have *J* and/or *Z* unless other unreported genes are involved in hilum ring color expression. As expected, the 15 darkening genotypes that had yellow hilum rings were Mayocobas. Mayocobas carry *gy* gene, which expresses yellow hilum ring regardless of *J* genotype (M. J. Bassett, 2002). Among the 48 genotypes that are non or slow darkening and had light/yellow corona (**Table S2.8**). All the 11 Mantecas were among the non or slow darkening genotypes with dark corona, and all the six beige genotypes were among the darkening genotypes with light/yellow corona. Genotypes in other market classes had different corona color and postharvest darkening behavior (**Table 2.4**). Mayocobas with yellow corona must be carrying *gy*, which expresses greenish-yellow corona color (M. J. Bassett, 2002). The 11 genotypes that had yellow hilum ring and the dark corona (**Table S2.9**) were all Mantecas, showing the uniqueness of this market class. 'Prim' Manteca had v^{lae}, which gives dark corona (M. J. Bassett, 1999); thus, the dark corona of the 11 Manteca genotypes is likely due to v^{lae}.

Genotype × Environment effects

The effects of genotype, environment, year, and interaction effects for the L*a*b* values are shown in **Table 2.5**. The genotype effect was significant for all the color values and postharvest darkening. The environment, genotype × environment, and genotype × environment × year effects for L*, a*, and b* values were significant, indicating that there is a genotype × environment interaction for the color values. The genotype × year was significant for a*, suggesting that a* of the seed colors in the YBC varied between years. However, the variation of a* due to year seems minimal because the ranges of a* were almost the same across the years and environments (**Table**

2.2) and the a* of the genotypes grown in each year were highly correlated ($R \geq 0.97$). The heritability estimates of the color values were high ($R \geq 0.93$, **Table 2.5**). This finding is consistent with other studies that reported high heritability of L*a*b* values with populations that had yellow or carioca beans as parental lines ($H^2 > 0.73$; Arns et al., 2018; A. Bassett et al., 2021; Possobom et al., 2015; Ribeiro et al., 2019). High heritability estimates of the color values indicate that genetic variance predominates in total phenotypic variance; thus, gain from selection is expected to be high.

Principal component analysis

Best linear unbiased estimation of color values adjusted for environment and year effects were used for principal component analysis (**Figure 2.3**). PC1 separated brown and Amarillo dark beans from the rest of the market classes explaining more than 60% of the total variance. PC2 separated genotypes based on b* values, explaining 33% of the total variance. PC1 and PC2 collectively explained 94% of the total variance, indicating that the color values of the YBC could be compressed to 2 dimensions instead of 3 without losing much information. This seemed to be because of the high negative correlation between L* and a* values (**Figure S2.7**). L* and a* are correlated mainly because of the Amarillo dark genotypes having higher a* and lower L* values than the other market classes of the YBC. Amarillo light, Canary, and Green-yellow genotypes overlapped with one another in the biplot (**Figure 2.3**), showing the similarity of seed colors of these color types, which corresponds to our observation. Manteca and Mayocoba beans formed relatively distinct clusters characterized by higher L* values than other seed types. Overall, the large and continuous variability of color values, even within market classes, indicated that the YBC is a source of various yellow colors that can meet specific consumer preferences.

Color differences between environments

Although the heritability for the L*a*b* values were high, genotype × environment effect was significant (**Table 2.5**); therefore, the seed colors of MI- and NE-grown seed were compared via ΔL*, Δa*, and Δb*, such that positive values indicate higher color values of the MI-grown seeds. The total differences were calculated as ΔE, a positive value that measures the degree of total color difference. The difference between two colors with ΔE of three or larger will be noticeable to the human eye (Haeghen et al., 2000). Many of the genotypes had ΔE larger than three (**Figure 2.4A**), indicating that their MI- and NE-grown seeds had noticeably different colors. ΔL* was negative for almost all the genotypes regardless of the seed types, indicating that MI-grown seeds were darker than NE-grown (**Figure 2.4B**). The absolute values of ΔL* were larger for some of the dark-colored seeds such as brown and Amarillo dark types, so they were even darker in MI as compared to NE than the lighter-colored yellow classes such as beige and Amarillo light. It agreed with previous literature and our observation that beans grown under humid conditions tend to produce darker seeds than those grown under dry weather (Osorno et al., 2018; Possobom et al., 2015). Higher seed coat lightness (L*) was also reported when a black × carioca population was grown in a dry season than in a rainy season (Possobom et al., 2015). Almost all the Amarillo light, Beige, Canary, Manteca, and Mayocoba beans had positive Δa* values, indicating that the seeds produced in MI were redder (**Figure 2.4C**). Amarillo dark and brown beans had lower Δb* values, meaning that they were less yellow in MI than NE (**Figure 2.4D**). Mayocoba and Canary had varying Δb*. Light colored beans such as beige and Manteca types had higher b* values, thus yellower in MI. Some Mayocoba type beans tend to produce paler colors under cool and wet fall in the Midwest US (Kelly et al., 2021). Indeed, 17 Mayocobas had negative Δb*, meaning that MI-grown beans were less yellow than NE. The other 33 Mayocobas, however, had positive Δb*,

which means yellower seeds were produced in MI, highlighting the importance of evaluating the color expression of Mayocoba beans in their target environments. Various factors could cause the difference in seed colors, such as humidity, rain, shorter exposure to sunlight, and increased pressure of certain diseases that favor moist conditions. Postharvest darkening was associated with the decrease in polymerization of proanthocyanidins in a regular darkening (RD) pinto bean (Beninger et al., 2005) and is characterized by the presence of proanthocyanidins and their precursors in RD cranberry beans (Chen et al., 2015), but further studies are needed to comprehensively understand the relationships between growing conditions, seed color, and polyphenolic profiles.

It was hypothesized that the genotypes that had darker seeds in MI (lower ΔL*, Δa*, or Δb* values) might be darkening types if darkening was already initiated before the images were taken. However, no clear differentiation in the color difference between darkening and non or slow darkening YBC lines was observed except for the b* values of Canary (**Figure S2.8**). Canary beans with higher yellowness in MI were darkening, whereas non or slow darkening Canary beans had lower yellowness in MI. Flavonoids, kaempferols in specific, are present in the seed coat of Canary beans (Hart et al., 2020), and it is possible that kaempferols present in darkening Canary beans grown in MI had started forming kaempferol-catechin adduct, which increased in RD pinto after darkening (Beninger et al., 2005).

Genome-wide association analysis

The Manhattan and QQ plots for the L*a*b* values, postharvest darkening, and hilum ring and corona colors are shown in **Figure 2.5**. Ten SNPs were significantly associated with a* value, nine with b*, 18 with L*, 5 for postharvest darkening, 9 for hilum ring color, and 7 for corona color

(**Table S2.10**). Three SNPs were significant for multiple traits: Chr02pos23133217 for a* and L*; Chr07pos29169848 for a*, L*, postharvest darkening, hilum ring color, and corona colors, and Chr10pos43341167 for a* and L*. Several SNPs were in proximity. Chr01pos41574533 for postharvest darkening and Chr01pos42211164 for b* was 637 kb apart. Chr02pos44319551 for L* and Chr02pos44931551 for L* were 612 kb apart. Chr02pos45646751 for hilum ring color, Chr02pos45663695 for a*, and Chr02pos46287738.1 for L* were all within the range of a 641-kb region. Chr05pos2078923 for hilum ring color and Chr05pos2408324 for corona color were 329 kb apart. Chr09pos18409319 for b* and Chr09pos18733021 for a* were 324 kb apart. Chr10pos42388055 for L* and Chr10pos42797440 for hilum ring color were 409 kb apart, and the latter was 544 kb apart from Chr10pos43341167 for a* and L*. Chr11pos4985949 for postharvest darkening and Chr11pos5038396 for hilum ring color were 52 kb apart. The phenotypic effect was larger for the SNPs for L* than a* and b*, reflecting the wider distribution of L* (**Table 2.1**).

Chr01pos50086405 on Pv01 was in a region where disease resistance genes are clustered, specifically for *Colletotrichum lindemuthianum* (Kelly & Bornowski, 2018). Polyphenolics play an important part in plant defense mechanisms against pests and pathogens (F. M. A. Islam et al., 2004; Bennett & Wallsgrove, 1994). If a gene is involved in the expression of seed colors important for consumers and disease resistance important for growers, the breeders' job would be to select genotypes that carry favorable alleles for both traits. It could be challenging and merit further research because an instance is reported that purple seed coat color and bean common mosaic virus resistance on Pv02 are tightly linked (or caused by a pleiotropic effect) and that it was not possible to introduce the resistance to other color types (Temple & Morales, 1986). Despite being beneficial

in insect control, polyphenols may adversely affect iron bioavailability (Hart et al., 2017); therefore, breeders need to balance the benefits and tradeoffs.

Some of the SNPs were supported by previous studies on seed coat color of pale lemon-yellow × green-yellow bean RIL population (A. Bassett et al., 2021): Chr03pos34038425 for L* on Pv03 was within the range of Sa*.3.1 (3.16-48.92 Mbp), Sb*.3.1 (29.62-41.84 Mbp), and SL*.3.1 (3.91-48.14 Mbp); Chr04pos1909564 for b* was within the range of Sb*.4.1 (0.13-2.06 Mbp); Chr10pos36447913, Chr10pos40877199, and Chr10pos42388055 for L* and Chr10pos43341167 for a* and L* were all within the range of Sa*10.1 (3.45-43.85 Mbp), Sb*.10.1 (37.83-43.47 Mbp), SL*.10.1 (3.45 - 43.85 Mbp), and ND.10.1 (3.98-43.85 Mbp); and Chr10pos42797440 for hilum ring color was within the range of the QTL for non-darkening, ND.10.1 (3.98 - 43.85 Mbp). Additionally, several SNPs coincided with QTL for quantitatively measured L*a*b* values of canned seeds detected in a black bean population (Bornowski et al., 2020): Chr10pos42388055 for L* was within the range of a*10.1BB (I) (42.23-43.30 Mbp), and Chr11pos53315933 for postharvest darkening was close to L*11.1 BB (I) (52.49–52.85 Mbp). Detection of multiple SNPs for multiple color values increases confidence in the association. Thus, these SNPs will be a useful reference in future studies on yellow color expression in common bean.

Chr07pos29169848 was significant for L*, a*, postharvest darkening, hilum ring color, and corona color with low *p*-values and large phenotypic effects (-10.0 for L*, +9.2 for a*, +0.59 for postharvest darkening, +0.90 for hilum ring color, and +0.64 for corona color; **Table S2.10**). The minor allele group genotypes at this SNP had a larger L*, a lower a*, a higher percentage of genotypes resistant to darkening, and light-colored hilum ring and corona (**Table S2.10**). Intriguingly, all the minor allele group genotypes (47 genotypes for L*, a*, b*, hilum ring color, and corona color, 45 for postharvest darkening) were Mayocoba at this SNP. Among them, 22

were from U.S. or Canada, and another 22 were from CIAT Colombia, whereas Mayocoba genotypes from other parts of the world such as Africa, Europe, and Mexico were found in the major allele group. Since the breeding materials from CIAT and those from the U.S. are closely related, this SNP may indicate the importance of color for the selection of Mayocoba type. This SNP is located near the ground factor *P* gene; the dominant *P* allele is necessary for seed coat color expression by other color genes (M. J. Bassett & Miklas, 2007; McClean, 2002; McClean et al., 2018). A QTL study with a Mayocoba × white bean population may reveal the potential role of the *P* gene (or another gene tightly linked to *P*) in giving the greenish yellow color in some of the Mayocoba beans. This SNP was also significant for postharvest darkening, and 36 out of the 45 Mayocoba genotypes in the minor allele group at this SNP were non or slow darkening. p^{sd}, an allele of *P*, is involved in slow darkening in the presence of *J*; genotypes with homozygous recessive p^{sd} shows a slower rate of darkening as compared to regular darkening genotypes (Elsadr et al., 2011; N. S. Islam et al., 2020; Junk-Knievel et al., 2008). The postharvest darkening behavior of the YBC genotypes, their SNP genotype at Chr07pos29169848, and their origin are summarized in **Table S2.6**. Surprisingly, 10 Canaries were non or slow darkening although Canaries were reported to have proanthocyanidins (Hart et al. 2020). They had the major allele at this SNP same as Mantecas, which do not have proanthocyanidins and are considered to be carrying *j* (Hart et al. 2020; Beninger et al. 1998). The difference between the non or slow darkening Canaries and Mantecas may be related to the other significant SNPs associated with postharvest darkening on Pv01, Pv04, and Pv11 (**Table S2.10**).

The Chr08pos3300438 associated with corona color on Pv08 was 69 kb apart from the *Gy* gene. *Gy* gene is considered to be responsible for the greenish-yellow seed coat, hilum ring, and corona colors of Mayocoba type (M. J. Bassett, 2002). However, the minor allele group of this

SNP included 11 Amarillos, 2 beige, and 1 white types which had light/yellow corona, and Mayocobas were all in the major allele group. The role of the *gy* gene in non-Mayocoba types is yet to be researched. None of the significant SNPs were close to known color genes such as *Z*, *G*, *B*, and *V* that interactively impart yellow to brown seed coat, hilum ring and/or corona colors (M. J. Bassett, 2007). There is no preceding study that quantitatively measured yellow seed color besides A. Bassett et al., (2021), thus further investigation will be necessary on genes and their interactions that produce the wide diversity of yellow colors.

The region of the *J* gene for brown hilum ring color and postharvest darkening has been narrowed down to 41,141,057 - 41,591,985 bp on Pv10 (Erfatpour et al., 2018), but Chr10pos42797440 associated with hilum ring color was approximately 1.2 Mbp away from that region. A higher marker density and more SNPs closer to the *J* gene would help investigate the hilum ring color and postharvest darkening of yellow beans. *J* was detected in this study for postharvest darkening but just below the significance threshold: Chr10pos42038959 on Pv10 near *J* (447 bp apart) had an FDR-adjusted *p*-value of 0.083. Nevertheless, a SNP (Chr10pos42797440) associated with hilum ring color within a range of a previously found QTL for postharvest darkening, ND.10.1 (3.98-43.85 Mbp; A. Bassett et al., 2021), lends support to the role of this region containing *J*. Another consideration for hilum ring color is that *J* and other genes are interacting – *J* is involved in the expression of dark hilum ring by v^{lae} with the absence of Z^{chr} (M. J. Bassett et al., 1999; M. J. Bassett, 2003), thus, hilum ring color may not be only determined by the genotype at *J*. Interestingly, Chr10pos40877199 on Pv10 associated with L* was close to the *J* gene (264 kb apart). *J* gene is involved in matured seed coat color development and postharvest darkening by regulating the biosynthesis pathways of polyphenolic compounds (M. J. Bassett, 2007; Elsadr et al., 2011; Erfatpour et al., 2018).

Conclusions

In this study, the seed coat color of yellow beans was measured by masking noise in bean images such as hilum ring, corona, background, and light reflection using machine-learning. As the trained model automatically segments seed coat from noise, it provides an efficient means to extract L*a*b* values from the seed coat. A large phenotypic diversity was observed for the L*a*b* values of the seed coat of YBC genotypes grown in MI and NE, USA. Despite the high heritability estimate for the color values, the genotype × environment effects were significant. Most of the YBC lines grown in MI and NE had a noticeable color difference, and those grown in MI with high humidity were darker. Genome-wide association analysis discovered SNPs associated with the color values, postharvest darkening, and hilum ring and corona colors of the YBC genotypes. The L*, a*, postharvest darkening, hilum ring color, and corona color mapped to the *P* (*sd*) gene. The phenotypic diversity and machine learning-aided image analysis of the seed coat color and other traits provide a resource to objectively evaluate and understand the seed morphology of beans, which will lead to the development of new varieties that better meet consumer's needs.

Acknowledgements

This work was supported by the USDA National Institute of Food and Agriculture AFRI (Award Number #: 2017-67013-26212), Michigan State University, the National Science Foundation Research Traineeship Program (DGE-1828149) awarded to Rie Sadohara and Yunfei Long, and the Nebraska Hatch Project (NEB43-116). The authors thank Anna Akariza, Amber Bassett, Gasana Ingabiregasana, Miranda Haus, Sharon Hooper, Queen Iribagiza, Hannah Jeffery, Hannah Peplinski, Scott Shaw, Jason Wiesinger, and Evan Wright for assistance with field operations and image acquisition.

APPENDICES

APPENDIX A:

CHAPTER 2 TABLES AND FIGURES

Table 2.1 Descriptive statistics of L*a*b* values of the YBC.

Trait	Environment-Year	Min	Max	Median	Mean	SD
L*	MI 2018	27.6	84.6	68.8	66.7	10.2
	MI 2019	31.0	84.5	70.7	67.8	9.7
	NE 2018	33.5	85.3	71.6	69.9	10.0
	NE 2019	31.9	86.0	73.1	71.7	9.8
a*	MI 2018	−4.7	30.6	8.0	9.4	8.0
	MI 2019	−3.5	32.1	8.4	9.9	8.0
	NE 2018	−5.9	33.2	7.2	8.2	9.5
	NE 2019	−5.6	35.0	5.8	7.4	9.7
b*	MI 2018	11.9	62.1	47.2	45.5	9.1
	MI 2019	12.6	62.2	47.8	46.9	9.0
	NE 2018	11.9	60.3	43.5	44.5	9.6
	NE 2019	12.3	63.6	45.5	45.7	10.1

Table 2.2 Darkening behavior of the YBC genotypes by seed type.

Seed type	Non or slow darkening	Darkening	Total
Amarillo (dark)	0	2	2
Amarillo (light)	4	38	42
Beige	0	6	6
Brown	0	3	3
Canary	10	64	74
Green-yellow	3	3	6
Manteca	11	0	11
Mayocoba	39	15	54
White	3	0	3
Amarillo dark and light mottled	0	1	1
Beige with brown stripes	0	3	3
White and red mottled	1	0	1
Total	71	135	206

Table 2.3 Hilum ring (HR) color of the YBC genotypes by seed type.

Seed type	Yellow HR	Dark HR	Total
Amarillo (dark)	0	36	36
Amarillo (light)	0	42	42
Beige	0	6	6
Brown	0	7	7
Canary	0	83	83
Green-yellow	0	6	6
Manteca	11	0	11
Mayocoba	56	0	56
White	3	0	3
Amarillo dark and light mottled	0	1	1
Amarillo (dark) with brown mottles	0	1	1
Amarillo (dark) with brown stripes	0	2	2
Beige with brown stripes	0	3	3
White and red mottled	0	1	1
Total	70	188	258

Table 2.4 Corona color of the YBC genotypes by seed type.

Seed type	Light/Yellow corona	Dark corona	Total
Amarillo (dark)	17	19	36
Amarillo (light)	28	14	42
Beige	6	0	6
Brown	1	6	7
Canary	0	83	83
Green-yellow	3	3	6
Manteca	0	11	11
Mayocoba	56	0	56
White	3	0	3
Amarillo dark and light mottled	1	0	1
Amarillo (dark) with brown mottles	1	0	1
Amarillo (dark) with brown stripes	2	0	2
Beige with brown stripes	3	0	3
White and red mottled	1	0	1
Total	122	136	258

Table 2.5 The *p*-values for the factor effects on L*a*b* values and postharvest darkening traits.

Term	L*	a*	b*	Darkening[a]
Genotype (G)	<0.0001	<0.0001	<0.0001	<0.0001
Environment (E)	<0.0001	<0.0001	0.046	.
Year (Y)	<0.0001	0.086	<0.0001	.
G × E	<0.0001	<0.0001	<0.0001	.
G × Y	0.42	0.015	1	.
E × Y	0.026	<0.0001	0.84	.
G × E × Y	<0.0001	<0.0001	<0.0001	.
[b]H^2	0.99	0.98	0.93	.

[a]Postharvest darkening was measured in MI 2019.

[b]Broad-sense heritability.

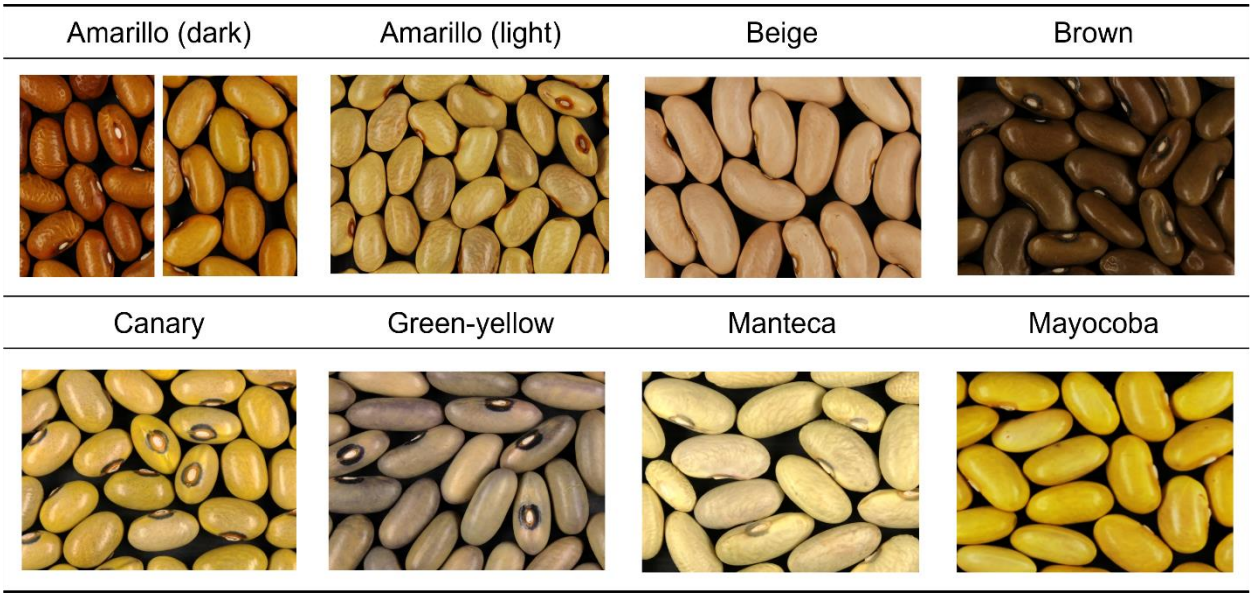


Figure 2.1 Eight major seed types in the YBC.

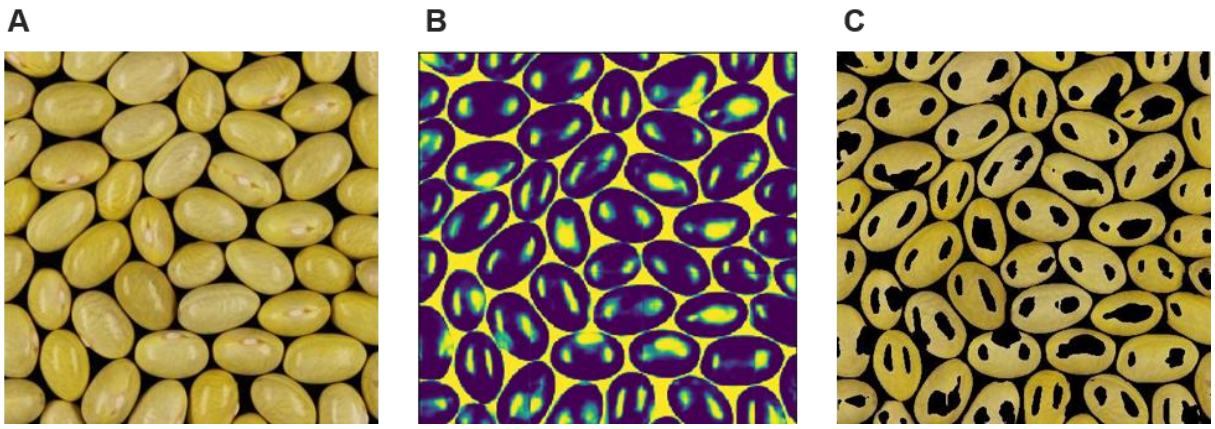


Figure 2.2 Original (**A**), segmented by convolutional neural networks (**B**), and masked (**C**) images. The probability of each pixel of being seed coat is color-coded from dark blue (high) to yellow (low) in **B**.

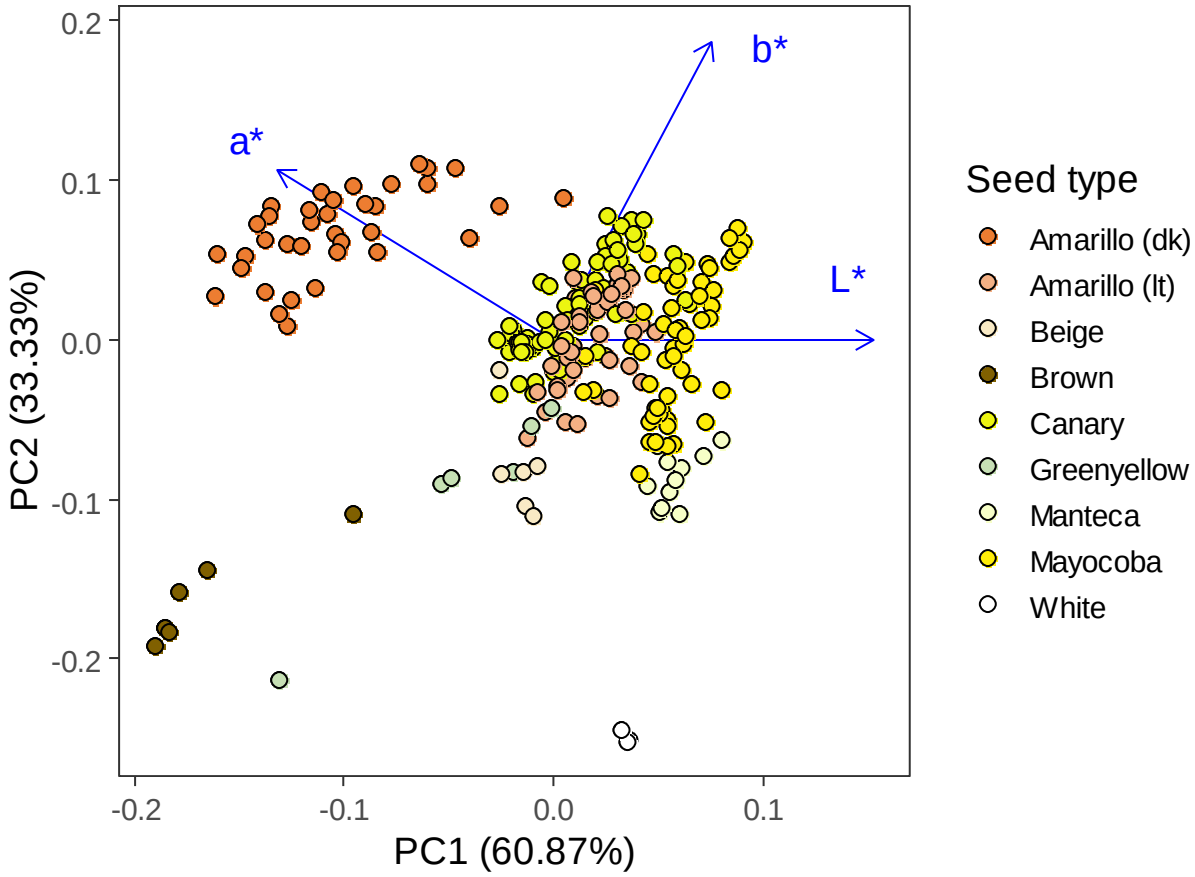


Figure 2.3 Biplot of principal component analysis of the L*a*b* values of the seed coat of YBC genotypes.

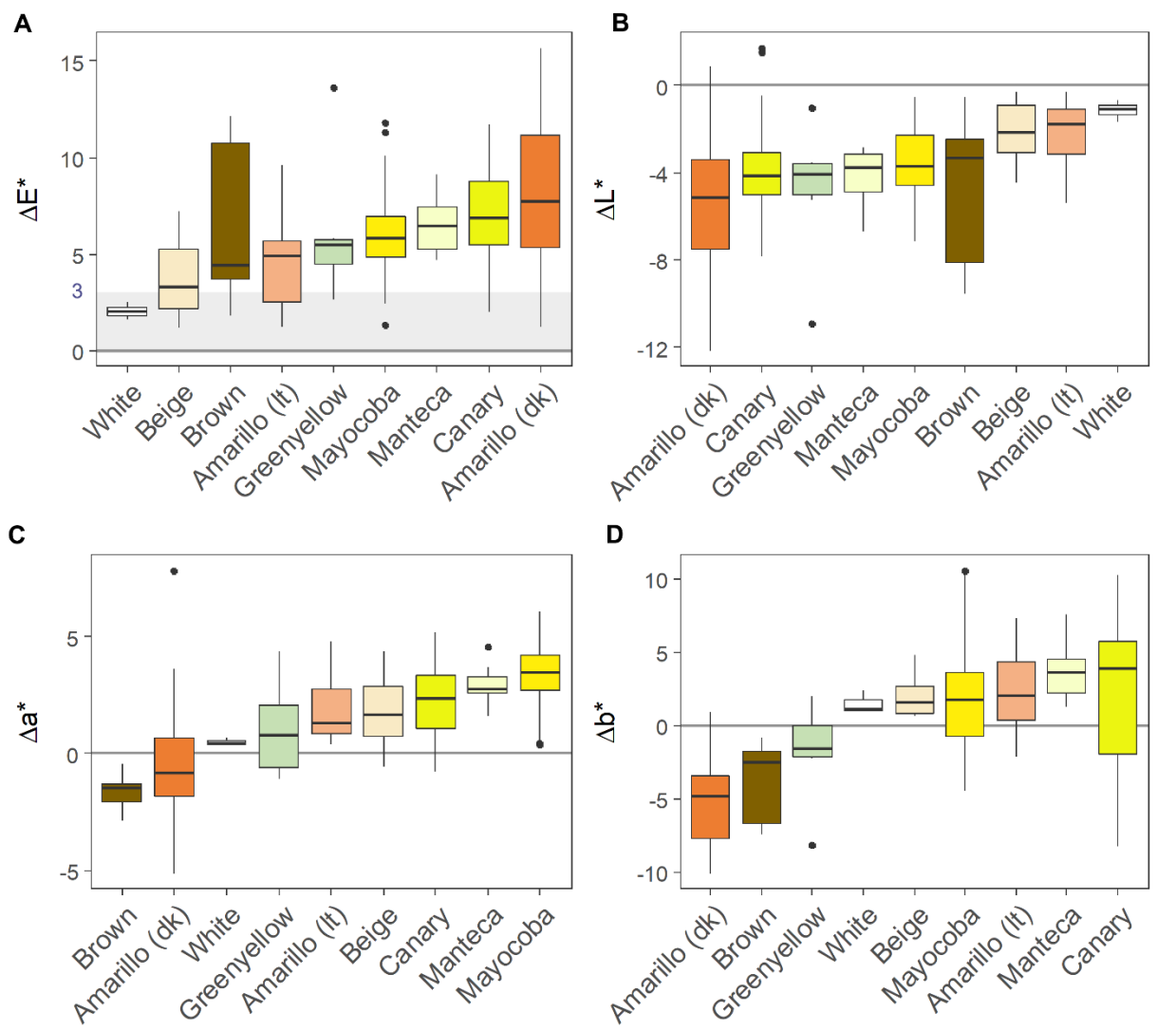


Figure 2.4 Color differences between MI and NE by seed type computed as differences between two-year average of L*a*b* values in MI and NE. **A**: ΔE*, **B**: ΔL*, **C**: Δa*, and **D**: Δb*.

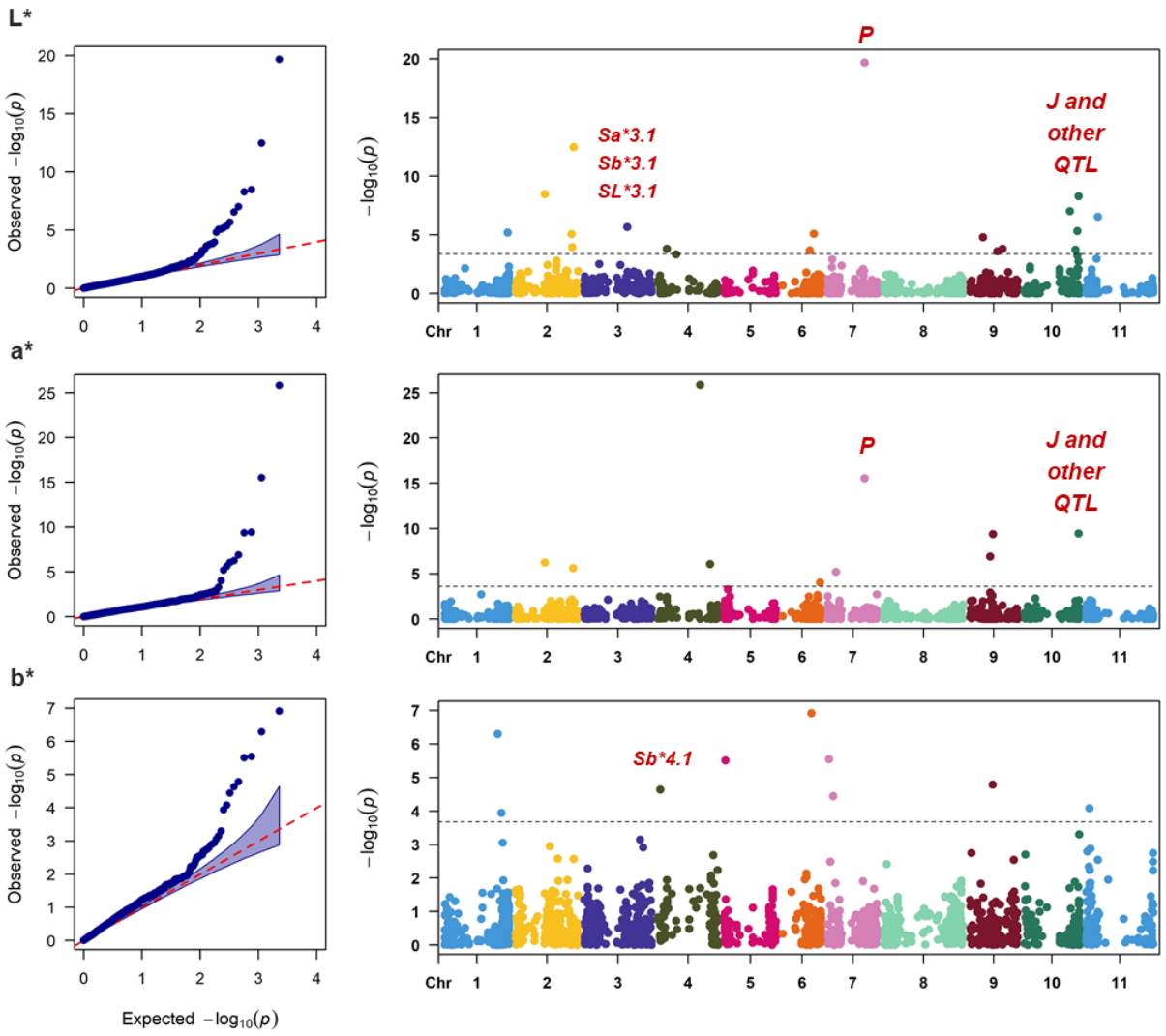
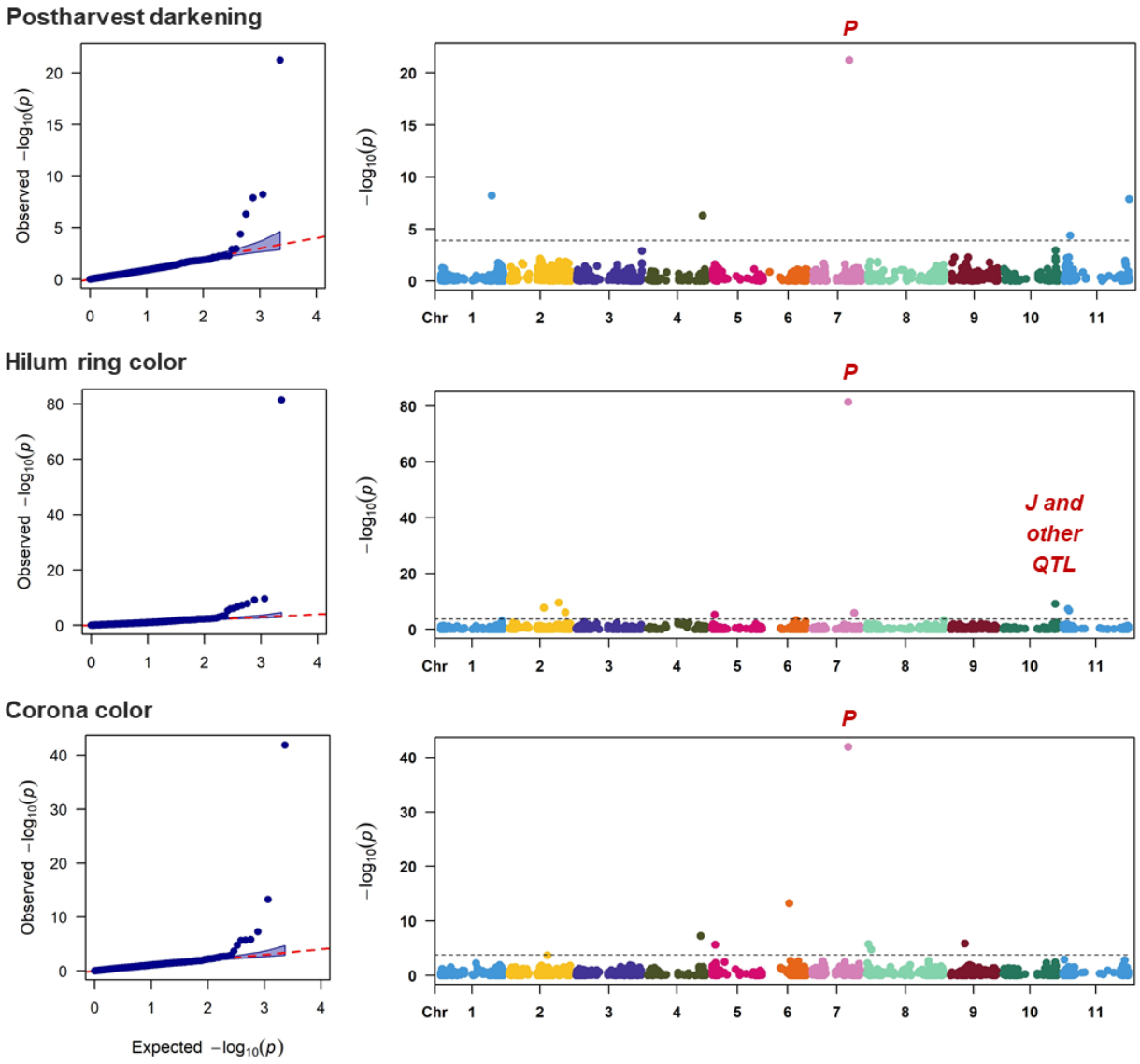


Figure 2.5 QQ and Manhattan plots for L*, a*, b*, postharvest darkening, hilum ring color, and corona color. The gray dotted lines are false discovery rate-adjusted threshold at α=0.05. *P:* the ground factor gene for seed coat color, of which p^{sd} allele confers slow darkening trait (N. S. Islam et al., 2020; McClean et al., 2018). *Sb*.4.1*, *Sa*.3.1*, *Sb*.3.1*, *SL*.3.1:* QTL reported by (A. Bassett et al., 2021). *J* and other QTL on Pv10: All the significant SNPs for L* are within the ranges of Sa*.10.1, Sb*.10.1, SL*.10.1, and ND.10.1 (A. Bassett et al., 2021). Chr10pos42388055 was within the range of L*10.1BB (I) (Bornowski et al., 2020) and *J* is the gene responsible for postharvest darkening (Elsadr et al., 2011). Chr10pos42388055 was within the range of a*10.1BB (I) (Bornowski et al., 2020).

Figure 2.5 (cont'd)



APPENDIX B:

CHAPTER 2 SUPPLEMENTAL TABLES AND FIGURES

Table S2.6 The number of genotypes phenotyped for color values and postharvest darkening.

	Environment × Year	Total
L*, a*, b*	MI 2018	252
	MI 2019	250
	NE 2018	209
	NE 2019	195
Postharvest darkening	MI 2019	206
Hilum ring and corona colors	MI 2019	258

Table S2.7 Number of genotypes in the postharvest darkening and hilum ring color categories.

		Hilum ring		
		Yellow	Dark	Total
Postharvest darkening	Non or slow - darkening	53	18	71
	Darkening	15	120	135
	Total	68	138	206

Table S2.8 Number of genotypes in the postharvest darkening and corona color categories.

		Corona		
		Light/yellow	Dark	Total
Postharvest darkening	Non or slow - darkening	48	23	71
	Darkening	53	82	135
	Total	101	105	206

Table S2.9 Number of genotypes in the hilum ring and corona color categories

		Corona		
		Light/yellow	Dark	Total
Hilum ring	Yellow	59	11	70
	Dark	63	125	188
	Total	122	136	258

Table S2.10 SNPs significantly associated with L*, a*, b*, postharvest darkening, hilum ring color, and corona color.

				------ Major allele ------			------ Minor allele ------				
SNP	Chr.	Position	FDR-adjusted *p*-values	Allele	n	Mean phenotypic value	Allele	n	Mean phenotypic value	MAF	SNP effect[a]
L*											
Chr01pos50086405	1	50,086,405	1.6E-03	AA	230	70.0	TT	14	60.5	0.057	9.5
Chr02pos23133217	2	23,133,217	2.5E-06	AA	173	68.7	GG	59	71.1	0.254	-2.3
Chr02pos44319551	2	44,319,551	1.8E-03	CC	164	69.8	TT	58	68.7	0.261	1.0
Chr02pos44931551	2	44,931,551	1.9E-02	GG	164	68.6	TT	62	70.4	0.274	-1.8
Chr02pos46287738	2	46,287,738	3.9E-10	GG	240	70.1	TT	13	55.2	0.051	14.9
Chr03pos34038425	3	34,038,425	6.9E-04	TT	207	69.0	CC	11	70.9	0.050	-1.9
Chr04pos7036250	4	7,036,250	2.3E-02	AA	173	68.8	TT	67	71.4	0.279	-2.6
Chr06pos21948635	6	21,948,635	2.8E-02	TT	187	68.9	CC	31	73.3	0.142	-4.4
Chr06pos24955965	6	24,955,965	1.8E-03	CC	126	67.6	TT	81	72.7	0.391	-5.1
Chr07pos29169848	7	29,169,848	4.8E-17	TT	200	67.3	GG	47	77.3	0.190	-10.0
Chr09pos10759327	9	10,759,327	2.9E-03	GG	231	68.9	AA	13	75.2	0.053	-6.3
Chr09pos22109781	9	22,109,781	3.1E-02	GG	184	68.8	AA	42	72.4	0.186	-3.7
Chr09pos26420953	9	26,420,953	2.3E-02	TT	144	69.5	GG	59	70.4	0.291	-0.9
Chr10pos36447913	10	36,447,913	4.3E-05	AA	179	69.8	GG	37	69.6	0.171	0.2
Chr10pos40877199	10	40,877,199	2.5E-02	AA	144	68.4	GG	70	71.0	0.327	-2.6
Chr10pos42388055	10	42,388,055	1.3E-03	CC	195	70.4	AA	19	66.1	0.089	4.3
Chr10pos43341167	10	43,341,167	2.9E-06	CC	167	70.6	TT	54	67.7	0.244	2.9
Chr11pos9547435	11	9,547,435	1.1E-04	CC	220	70.4	TT	15	57.5	0.064	13.0
a*											
Chr02pos23133217	2	23,133,217	2.2E-04	AA	173	8.5	GG	59	9.3	0.254	-0.8
Chr02pos45663695	2	45,663,695	6.8E-04	AA	133	8.8	TT	78	7.6	0.370	1.1
Chr04pos33747479	4	33,747,479	3.3E-23	GG	153	5.2	AA	90	14.2	0.370	-9.1
Chr04pos41620411	4	41,620,411	2.9E-04	CC	155	8.2	TT	57	8.9	0.269	-0.7
Chr06pos30013548	6	30,013,548	2.2E-02	GG	194	8.5	AA	53	8.9	0.215	-0.4
Chr07pos6344071	7	6,344,071	1.6E-03	AA	157	8.5	GG	50	8.4	0.242	0.1

Table S2.10 (cont'd)

Chr07pos29169848	7	29,169,848	3.4E-13	TT	200	10.5	GG	47	1.3	0.190	9.2
Chr09pos16476454	9	16,476,454	6.0E-05	TT	166	7.5	CC	51	10.3	0.235	-2.8
Chr09pos18733021	9	18,733,021	2.5E-07	TT	158	9.6	AA	54	7.8	0.255	1.8
Chr10pos43341167	10	43,341,167	2.5E-07	CC	167	6.8	TT	54	13.4	0.244	-6.6
b*											
Chr01pos42211164	1	42,211,164	5.8E-04	GG	167	45.0	AA	44	47.7	0.209	-2.7
Chr01pos45147107	1	45,147,107	2.9E-02	CC	190	46.9	TT	17	40.0	0.082	6.9
Chr04pos1909564	4	1,909,564	8.8E-03	TT	201	47.3	AA	28	40.5	0.122	6.8
Chr05pos817906	5	817,906	1.8E-03	GG	123	48.4	AA	97	43.4	0.441	5.0
Chr06pos23037583	6	23,037,583	2.8E-04	GG	203	44.9	AA	32	51.5	0.136	-6.6
Chr07pos962487	7	962,487	1.8E-03	CC	200	45.5	AA	21	49.1	0.095	-3.6
Chr07pos4181998	7	4,181,998	1.2E-02	TT	187	46.2	CC	33	43.7	0.150	2.6
Chr09pos18409319	9	18,409,319	7.5E-03	CC	214	45.9	TT	13	45.7	0.057	0.2
Chr11pos2684350	11	2,684,350	2.4E-02	GG	116	47.5	AA	114	44.8	0.496	2.7
Postharvest darkening											
Chr01pos41574533	1	41,574,533	6.7E-06	TT	192	0.68	AA	14	0.29	0.068	0.40
Chr04pos44665073	4	44,665,073	2.8E-04	CC	165	0.68	GG	21	0.48	0.113	0.21
Chr07pos29169848	7	29,169,848	1.3E-18	TT	155	0.79	GG	45	0.20	0.225	0.59
Chr11pos4985949	11	4,985,949	2.0E-02	GG	181	0.67	TT	16	0.44	0.081	0.24
Chr11pos53315933	11	53,315,933	9.9E-06	TT	133	0.56	CC	37	0.89	0.218	-0.34
Hilum ring color											
Chr02pos28089361	2	28,089,361	1.0E-05	TT	199	0.73	CC	28	0.71	0.123	0.01
Chr02pos40280426	2	40,280,426	2.8E-07	TT	172	0.70	CC	62	0.77	0.265	-0.08
Chr02pos45646751	2	45,646,751	2.6E-04	TT	169	0.67	AA	73	0.86	0.302	-0.19
Chr05pos2078923	5	2,078,923	1.4E-03	AA	164	0.67	GG	48	0.90	0.226	-0.23
Chr07pos29169848	7	29,169,848	9.1E-79	TT	205	0.90	GG	47	0.00	0.187	0.90
Chr07pos34201349	7	34,201,349	3.9E-04	CC	204	0.69	TT	11	1.00	0.051	-0.31
Chr10pos42797440	10	42,797,440	5.5E-07	AA	146	0.70	GG	75	0.79	0.339	-0.09
Chr11pos4112749	11	4,112,749	2.4E-05	TT	204	0.75	GG	16	0.38	0.073	0.38

Table S2.10 (cont'd)

Chr11pos5038396	11	5,038,396	7.6E-05	AA	226	0.73	GG	21	0.62	0.085	0.12
Corona color											
Chr04pos43635399	4	43,635,399	4.5E-05	GG	210	0.50	CC	14	0.79	0.063	-0.29
Chr05pos2408324	5	2,408,324	8.8E-04	AA	168	0.68	TT	40	0.23	0.192	0.45
Chr06pos17242131	6	17,242,131	6.9E-11	TT	200	0.60	AA	20	0.30	0.091	0.30
Chr07pos29169848	7	29,169,848	2.9E-39	TT	205	0.64	GG	47	0.00	0.187	0.64
Chr08pos1159204	8	1,159,204	8.6E-04	AA	196	0.54	GG	15	0.33	0.071	0.21
Chr08pos3300438	8	3,300,438	6.4E-03	GG	222	0.61	AA	14	0.00	0.059	0.61
Chr09pos12073563	9	12,073,563	8.6E-04	TT	190	0.64	CC	25	0.16	0.116	0.48

[a]SNP effect is the difference between the mean cooking times of the YBC lines with the major and minor alleles.

Table S2.11 BLUE Lab.xlsx. The best linear unbiased estimates (BLUE) of seed coat L*a*b* of the Yellow Bean Collection adjusted for environment and year effects.

Taxa	BLUE L*	BLUE a*	BLUE b*
YBC001	63.5	24.1	52.4
YBC002	68.1	10.3	44.6
YBC003	68.1	10.0	44.1
YBC004	84.8	-4.1	13.5
YBC005	53.3	32.0	45.7
YBC006	66.1	10.9	45.5
YBC007	68.3	9.7	40.8
YBC010	75.7	2.5	45.4
YBC011	51.4	23.7	38.5
YBC012	66.7	10.8	43.5
YBC013	70.4	8.1	36.4
YBC014	72.4	7.9	42.5
YBC015	74.3	7.7	38.3
YBC017	72.0	7.6	58.5
YBC018	48.6	32.1	40.8
YBC021	67.2	10.8	50.5
YBC022	67.8	10.3	50.3
YBC024	74.8	3.7	46.1
YBC025	68.5	9.8	44.4
YBC026	67.6	10.6	44.6
YBC028	73.6	6.1	39.0
YBC029	71.6	8.4	39.0
YBC030	76.0	5.1	49.6
YBC034	74.7	6.6	45.5
YBC035	76.9	4.1	49.3
YBC036	66.8	11.4	44.7
YBC041	73.8	5.8	56.1
YBC045	66.5	11.4	44.9
YBC046	67.8	10.3	45.1
YBC047	73.7	5.0	42.4
YBC049	80.6	-2.5	47.5
YBC050	52.8	31.3	45.0
YBC051	52.9	28.3	44.0
YBC052	67.4	10.6	44.0
YBC053	70.3	8.2	43.0
YBC054	72.1	6.3	57.9
YBC055	73.2	6.6	46.0
YBC056	67.7	7.6	42.3
YBC057	73.2	7.5	44.7
YBC059	60.7	25.9	51.4

Table S2.11 (cont'd)

YBC061	49.9	30.1	41.8
YBC063	80.7	-4.6	43.8
YBC064	70.7	7.5	46.1
YBC065	70.0	10.5	52.5
YBC066	73.4	5.6	42.8
YBC067	70.0	7.8	44.1
YBC068	71.5	8.0	48.5
YBC069	69.8	7.2	43.7
YBC070	69.8	8.6	45.2
YBC071	76.9	-1.0	41.0
YBC072	67.1	10.8	44.4
YBC073	67.7	10.1	44.5
YBC075	46.6	29.0	38.4
YBC076	78.4	-3.1	40.1
YBC077	50.6	27.0	40.0
YBC078	65.9	23.6	54.3
YBC079	69.4	7.9	43.8
YBC080	68.1	9.9	44.3
YBC081	67.2	10.6	44.5
YBC082	69.5	9.6	46.0
YBC083	67.6	10.7	44.5
YBC084	67.8	10.3	44.4
YBC085	67.9	10.1	44.2
YBC086	67.7	10.3	44.1
YBC087	74.9	5.0	54.6
YBC088	67.6	9.9	44.7
YBC089	70.2	8.1	46.3
YBC090	71.0	6.6	45.5
YBC091	64.6	6.8	49.0
YBC092	67.3	10.7	44.4
YBC093	67.5	10.4	43.9
YBC094	64.3	25.5	53.2
YBC095	68.6	9.5	45.2
YBC096	84.9	-4.0	13.3
YBC097	52.2	25.1	40.2
YBC098	70.5	8.1	55.8
YBC099	56.9	28.5	46.9
YBC100	71.8	9.8	50.6
YBC101	77.4	-2.3	40.1
YBC103	80.0	-3.6	35.8
YBC105	78.4	-3.0	37.6
YBC106	71.6	9.8	49.4

Table S2.11 (cont'd)

YBC107	68.7	8.0	40.8
YBC108	67.7	10.7	44.6
YBC109	72.5	6.6	59.0
YBC111	79.7	-2.8	44.5
YBC112	67.6	9.7	44.0
YBC113	51.6	29.9	43.4
YBC114	71.7	7.9	55.7
YBC115	67.5	18.7	53.4
YBC116	76.5	0.5	42.4
YBC117	66.8	6.2	50.2
YBC118	67.2	6.9	52.7
YBC119	70.9	7.8	54.8
YBC120	66.7	6.4	53.0
YBC121	74.7	4.7	49.0
YBC122	77.7	1.8	55.7
YBC123	76.7	1.2	49.4
YBC124	79.3	-2.1	35.1
YBC125	70.9	7.6	58.0
YBC126	78.4	0.7	55.5
YBC127	75.3	1.3	45.7
YBC128	64.5	3.7	40.3
YBC129	78.7	-2.8	38.7
YBC130	62.0	9.1	31.4
YBC131	75.6	5.2	42.2
YBC132	63.0	25.4	53.7
YBC133	35.2	11.7	19.1
YBC134	38.0	11.9	21.1
YBC135	69.6	13.7	57.1
YBC136	70.5	9.1	58.0
YBC137	78.2	-0.6	36.7
YBC138	57.0	30.1	48.1
YBC140	53.7	27.4	44.3
YBC142	66.5	5.8	50.9
YBC143	67.7	7.0	52.8
YBC144	53.3	27.1	44.4
YBC145	32.9	9.8	16.6
YBC146	69.6	7.0	52.9
YBC147	71.4	5.6	52.9
YBC148	69.1	7.5	50.8
YBC149	67.1	5.7	34.4
YBC150	31.7	9.3	15.0
YBC151	32.9	9.9	16.6

Table S2.11 (cont'd)

YBC153	68.2	5.6	49.1
YBC154	71.7	5.5	54.5
YBC155	70.2	6.1	50.2
YBC157	71.9	6.0	56.9
YBC159	73.2	7.8	29.9
YBC160	69.9	9.1	32.3
YBC161	73.7	9.7	32.0
YBC162	72.3	7.8	51.6
YBC163	68.0	10.6	43.6
YBC165	73.2	6.5	29.6
YBC166	72.4	6.9	57.4
YBC167	38.2	-0.3	17.4
YBC168	57.0	26.8	45.8
YBC169	65.6	11.0	45.8
YBC170	66.0	3.4	42.2
YBC171	71.1	6.6	55.3
YBC172	70.4	14.1	39.8
YBC173	63.0	8.9	32.0
YBC175	71.3	7.8	41.8
YBC176	75.9	2.7	50.9
YBC177	77.2	1.8	51.2
YBC178	77.0	1.9	47.9
YBC179	65.8	6.0	50.8
YBC180	33.1	9.4	16.4
YBC181	51.4	8.8	28.4
YBC182	78.6	-2.6	35.7
YBC184	49.2	29.5	40.9
YBC186	79.6	-4.2	41.8
YBC190	70.4	8.0	54.0
YBC191	71.8	7.2	34.3
YBC192	70.0	8.7	40.7
YBC193	77.2	1.8	53.5
YBC194	78.1	2.3	57.3
YBC195	78.8	0.6	54.0
YBC196	79.2	0.8	58.2
YBC197	75.0	3.6	55.2
YBC198	77.5	1.0	49.9
YBC199	75.5	3.0	54.6
YBC200	83.6	-3.8	14.4
YBC203	56.8	28.3	48.4
YBC204	70.8	7.7	56.4
YBC206	76.6	1.7	51.0

Table S2.11 (cont'd)

YBC207	77.3	0.0	48.0
YBC208	79.4	0.7	58.8
YBC209	80.0	0.8	60.3
YBC210	79.4	1.6	61.1
YBC211	79.9	0.7	59.5
YBC212	79.0	1.8	60.1
YBC213	79.6	0.9	52.6
YBC214	76.2	3.3	56.8
YBC215	77.3	0.5	43.7
YBC216	76.7	0.8	43.7
YBC217	71.7	6.4	45.5
YBC221	74.4	4.1	47.9
YBC222	78.2	1.9	57.0
YBC223	73.3	6.3	42.1
YBC224	77.2	0.1	37.6
YBC225	74.8	3.7	57.4
YBC226	72.0	4.2	51.3
YBC227	77.7	-0.2	43.1
YBC228	77.8	-0.6	42.6
YBC229	71.7	6.9	54.8
YBC230	76.8	0.0	40.9
YBC231	76.9	0.2	43.6
YBC232	72.6	4.1	51.1
YBC233	77.6	0.6	45.0
YBC234	78.5	-1.4	41.4
YBC235	78.3	-1.0	40.9
YBC236	74.6	4.4	51.1
YBC237	76.2	3.3	55.1
YBC238	75.4	3.6	47.7
YBC239	76.7	0.4	51.1
YBC240	78.8	-0.5	46.8
YBC241	77.0	0.8	42.9
YBC242	77.6	0.0	40.7
YBC243	74.7	2.7	56.7
YBC244	77.0	2.2	53.6
YBC245	77.7	1.4	54.6
YBC246	78.0	0.4	52.6
YBC247	76.8	0.6	43.9
YBC248	76.9	1.1	48.7
YBC249	76.0	2.6	52.5
YBC250	56.2	25.3	46.0
YBC253	54.2	24.5	41.8

Table S2.11 (cont'd)

YBC254	71.3	7.2	54.3
YBC256	59.3	25.4	49.4
YBC257	65.4	11.6	44.3
YBC258	69.2	7.4	50.3
YBC259	65.8	6.8	47.0
YBC260	61.7	6.2	41.9
YBC262	66.4	5.7	49.7
YBC263	54.9	28.2	46.2
YBC265	70.4	6.4	56.2
YBC266	71.1	7.6	51.4
YBC268	55.4	29.4	46.8
YBC273	52.0	31.6	44.0
YBC274	59.3	28.7	49.6
YBC275	60.3	27.3	48.5
YBC278	50.8	24.6	39.2
YBC279	66.1	19.6	49.7
YBC280	55.9	25.1	45.1
YBC281	59.3	24.8	47.3
YBC282	72.3	8.6	44.0
YBC283	73.1	7.6	50.6
YBC284	74.4	7.2	51.8
YBC285	74.1	6.9	49.7
YBC286	74.3	6.7	53.1
YBC287	73.2	7.3	52.3
YBC288	72.7	7.9	49.3
YBC289	73.1	7.4	50.5
YBC290	72.7	7.3	53.3
YBC291	74.5	7.7	51.7
YBC292	74.1	7.5	52.0
YBC293	75.4	6.6	45.0
YBC294	70.8	10.4	51.1
YBC295	59.9	23.8	45.8
YBC296	73.8	8.9	42.8
YBC297	72.8	9.3	50.3
YBC298-1	73.5	7.6	51.2
YBC299	72.4	8.7	50.3
YBC300	71.6	8.6	48.3
YBC301	73.5	7.3	47.3
YBC302	72.7	8.9	40.8
YBC303	72.0	8.9	47.7
YBC304	72.9	6.1	50.0
YBC305	75.5	7.2	49.8

Table S2.11 (cont'd)

YBC306	71.8	10.5	46.5
YBC307	72.8	8.9	44.5
YBC308	72.6	10.2	44.4

Table S2.12 Postharvest darkening.xlsx. The postharvest darkening behavior of the Yellow Bean Collection.

Taxa	Postharvest darkening (0=non or slow darkening, 1=darkening)
YBC002	1
YBC003	1
YBC004	0
YBC006	1
YBC007	1
YBC010	1
YBC012	1
YBC014	1
YBC015	1
YBC017	0
YBC021	1
YBC022	1
YBC024	1
YBC025	1
YBC026	1
YBC028	1
YBC029	1
YBC030	0
YBC034	1
YBC035	0
YBC036	1
YBC041	1
YBC045	1
YBC046	1
YBC047	1
YBC049	0
YBC052	1
YBC053	1
YBC054	0
YBC055	1
YBC056	1
YBC057	1
YBC063	0
YBC064	1
YBC065	1
YBC066	1

Table S2.12 (cont'd)

YBC067	1
YBC068	1
YBC069	1
YBC070	1
YBC071	0
YBC072	1
YBC073	1
YBC076	0
YBC078	1
YBC079	1
YBC080	1
YBC081	1
YBC082	1
YBC083	1
YBC084	1
YBC085	1
YBC086	1
YBC087	1
YBC088	1
YBC089	1
YBC090	1
YBC091	1
YBC092	1
YBC093	1
YBC095	1
YBC096	0
YBC098	1
YBC100	1
YBC101	0
YBC102	0
YBC103	0
YBC105	0
YBC106	1
YBC107	1
YBC108	1
YBC109	0
YBC111	0
YBC112	1
YBC115	1
YBC116	0
YBC117	1
YBC118	0

Table S2.12 (cont'd)

YBC119	1
YBC121	0
YBC122	0
YBC123	0
YBC124	0
YBC125	0
YBC126	0
YBC127	0
YBC128	0
YBC129	0
YBC130	1
YBC131	1
YBC133	1
YBC134	1
YBC136	1
YBC137	0
YBC141	1
YBC143	0
YBC146	1
YBC147	1
YBC148	1
YBC149	1
YBC152	1
YBC153	1
YBC154	1
YBC155	1
YBC156	1
YBC157	1
YBC159	1
YBC160	1
YBC161	1
YBC162	1
YBC163	1
YBC165	1
YBC166	0
YBC167	0
YBC169	1
YBC170	0
YBC171	0
YBC172	1
YBC173	1
YBC174	1

Table S2.12 (cont'd)

YBC175	1
YBC176	0
YBC177	0
YBC178	0
YBC181	1
YBC182	0
YBC186	0
YBC191	1
YBC192	0
YBC193	0
YBC194	1
YBC195	0
YBC196	0
YBC197	0
YBC198	1
YBC199	0
YBC200	0
YBC204	0
YBC206	0
YBC207	0
YBC208	0
YBC209	0
YBC210	0
YBC211	0
YBC212	0
YBC214	0
YBC215	1
YBC217	1
YBC221	0
YBC222	1
YBC223	0
YBC224	0
YBC225	1
YBC226	1
YBC227	0
YBC228	0
YBC230	0
YBC231	0
YBC232	1
YBC233	0
YBC234	0
YBC235	0

Table S2.12 (cont'd)

YBC236	1
YBC237	0
YBC238	1
YBC239	0
YBC240	1
YBC241	1
YBC242	1
YBC243	0
YBC244	1
YBC245	0
YBC246	0
YBC247	0
YBC248	1
YBC249	0
YBC254	1
YBC257	1
YBC258	1
YBC265	1
YBC266	1
YBC282	1
YBC283	1
YBC284	1
YBC285	1
YBC286	1
YBC287	1
YBC288	1
YBC289	1
YBC291	1
YBC292	1
YBC293	1
YBC294	1
YBC296	1
YBC297	1
YBC298-1	1
YBC299	1
YBC300	1
YBC301	1
YBC302	1
YBC303	1
YBC304	1
YBC305	1
YBC306	1

Table S2.12 (cont'd)

YBC307	1
YBC308	1

Table S2.13 HR and corona colors.xlsx. The hilum ring and corona colors of the Yellow Bean Collection.

Taxa	Hilum ring color (0=Yellow, 1=Dark)	Corona color (0=Light/yellow, 2=Dark)
YBC001	1	0
YBC002	1	1
YBC003	1	1
YBC004	0	0
YBC005	1	1
YBC006	1	1
YBC007	1	1
YBC010	1	0
YBC011	1	0
YBC012	1	1
YBC014	1	0
YBC015	1	0
YBC017	1	1
YBC018	1	1
YBC021	1	1
YBC022	1	1
YBC024	1	0
YBC025	1	1
YBC026	1	1
YBC028	1	0
YBC029	1	0
YBC030	1	0
YBC034	1	0
YBC035	1	0
YBC036	1	1
YBC041	0	0
YBC045	1	1
YBC046	1	1
YBC047	1	0
YBC049	0	0
YBC050	1	1
YBC051	1	1
YBC052	1	1
YBC053	1	1
YBC054	1	1
YBC055	1	1
YBC056	1	1

Table S2.13 (cont'd)

YBC057	1	0
YBC059	1	0
YBC061	1	1
YBC063	0	1
YBC064	1	0
YBC065	1	1
YBC066	0	0
YBC067	1	1
YBC068	1	1
YBC069	1	1
YBC070	1	1
YBC071	0	1
YBC072	1	1
YBC073	1	1
YBC075	1	1
YBC076	0	1
YBC077	1	0
YBC078	1	0
YBC079	1	0
YBC080	1	1
YBC081	1	1
YBC082	1	1
YBC083	1	1
YBC084	1	1
YBC085	1	1
YBC086	1	1
YBC087	0	0
YBC088	1	1
YBC089	1	1
YBC090	1	1
YBC091	1	1
YBC092	1	1
YBC093	1	1
YBC094	1	0
YBC095	1	1
YBC096	0	0
YBC097	1	1
YBC098	1	1
YBC099	1	0
YBC100	1	1
YBC101	0	1
YBC102	1	0

Table S2.13 (cont'd)

YBC103	0	1
YBC105	0	1
YBC106	1	0
YBC107	1	1
YBC108	1	1
YBC109	1	1
YBC111	0	0
YBC112	1	1
YBC113	1	1
YBC114	1	1
YBC115	1	1
YBC116	0	0
YBC117	1	1
YBC118	1	1
YBC119	1	1
YBC120	1	1
YBC121	1	1
YBC122	0	0
YBC123	0	0
YBC124	0	1
YBC125	1	1
YBC126	0	0
YBC127	0	0
YBC128	1	0
YBC129	0	1
YBC130	1	1
YBC131	1	0
YBC132	1	0
YBC133	1	1
YBC134	1	1
YBC135	1	1
YBC136	1	1
YBC137	0	1
YBC138	1	0
YBC139	1	0
YBC140	1	1
YBC141	1	0
YBC142	1	1
YBC143	1	1
YBC144	1	1
YBC145	1	1
YBC146	1	1

Table S2.13 (cont'd)

YBC147	1	1
YBC148	1	1
YBC149	1	0
YBC150	1	1
YBC151	1	1
YBC152	1	0
YBC153	1	1
YBC154	1	1
YBC155	1	1
YBC156	1	0
YBC157	1	1
YBC159	1	0
YBC160	1	0
YBC161	1	0
YBC162	1	1
YBC163	1	1
YBC165	1	0
YBC166	1	1
YBC167	1	1
YBC168	1	1
YBC169	1	1
YBC170	1	0
YBC171	1	1
YBC172	1	0
YBC173	1	1
YBC174	1	0
YBC175	1	1
YBC176	0	0
YBC177	0	0
YBC178	0	0
YBC179	1	1
YBC180	1	1
YBC181	1	0
YBC182	0	1
YBC183	1	0
YBC184	1	0
YBC186	0	1
YBC187	1	0
YBC190	1	1
YBC191	1	0
YBC192	1	0
YBC193	0	0

Table S2.13 (cont'd)

YBC194	0	0
YBC195	0	0
YBC196	0	0
YBC197	0	0
YBC198	0	0
YBC199	0	0
YBC200	0	0
YBC203	1	1
YBC204	1	1
YBC206	0	0
YBC207	0	0
YBC208	0	0
YBC209	0	0
YBC210	0	0
YBC211	0	0
YBC212	0	0
YBC213	0	0
YBC214	0	0
YBC215	0	0
YBC216	0	0
YBC217	0	0
YBC221	0	0
YBC222	0	0
YBC223	0	0
YBC224	0	0
YBC225	1	1
YBC226	1	1
YBC227	0	0
YBC228	0	0
YBC229	1	1
YBC230	0	0
YBC231	0	0
YBC232	1	1
YBC233	0	0
YBC234	0	0
YBC235	0	0
YBC236	0	0
YBC237	0	0
YBC238	0	0
YBC239	0	0
YBC240	0	0
YBC241	0	0

Table S2.13 (cont'd)

YBC242	0	0
YBC243	1	1
YBC244	1	1
YBC245	0	0
YBC246	0	0
YBC247	0	0
YBC248	0	0
YBC249	0	0
YBC250	1	1
YBC253	1	1
YBC254	1	1
YBC256	1	1
YBC257	1	1
YBC258	1	1
YBC259	1	1
YBC260	1	1
YBC262	1	1
YBC263	1	1
YBC265	1	1
YBC266	0	0
YBC268	1	1
YBC274	1	0
YBC275	1	0
YBC278	1	0
YBC279	1	0
YBC280	1	0
YBC281	1	0
YBC282	1	0
YBC283	1	1
YBC284	1	0
YBC285	1	0
YBC286	1	1
YBC287	1	1
YBC288	1	1
YBC289	1	1
YBC291	1	1
YBC292	1	1
YBC293	1	0
YBC294	1	1
YBC295	1	0
YBC296	1	0
YBC297	1	1

Table S2.13 (cont'd)

YBC298-1	1	1
YBC299	1	0
YBC300	1	0
YBC301	1	1
YBC302	1	0
YBC303	1	1
YBC304	1	1
YBC305	1	0
YBC306	1	0
YBC307	1	0
YBC308	1	0

Figure S2.6 Examples of diversity in seed morphology of the YBC.

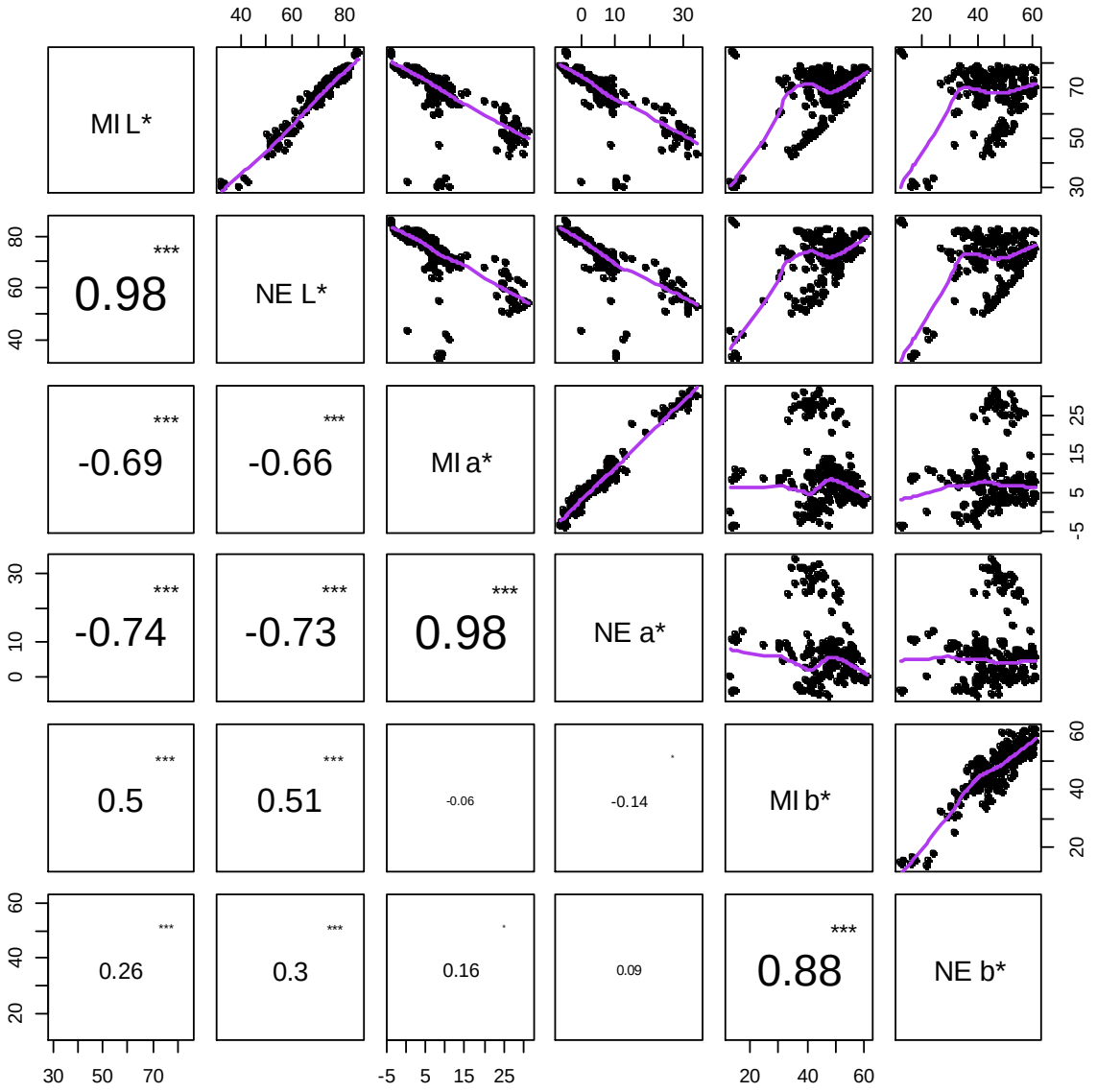


Figure S2.7 Pearson correlation between two-year average of L*a*b* in each environment (MI and NE).

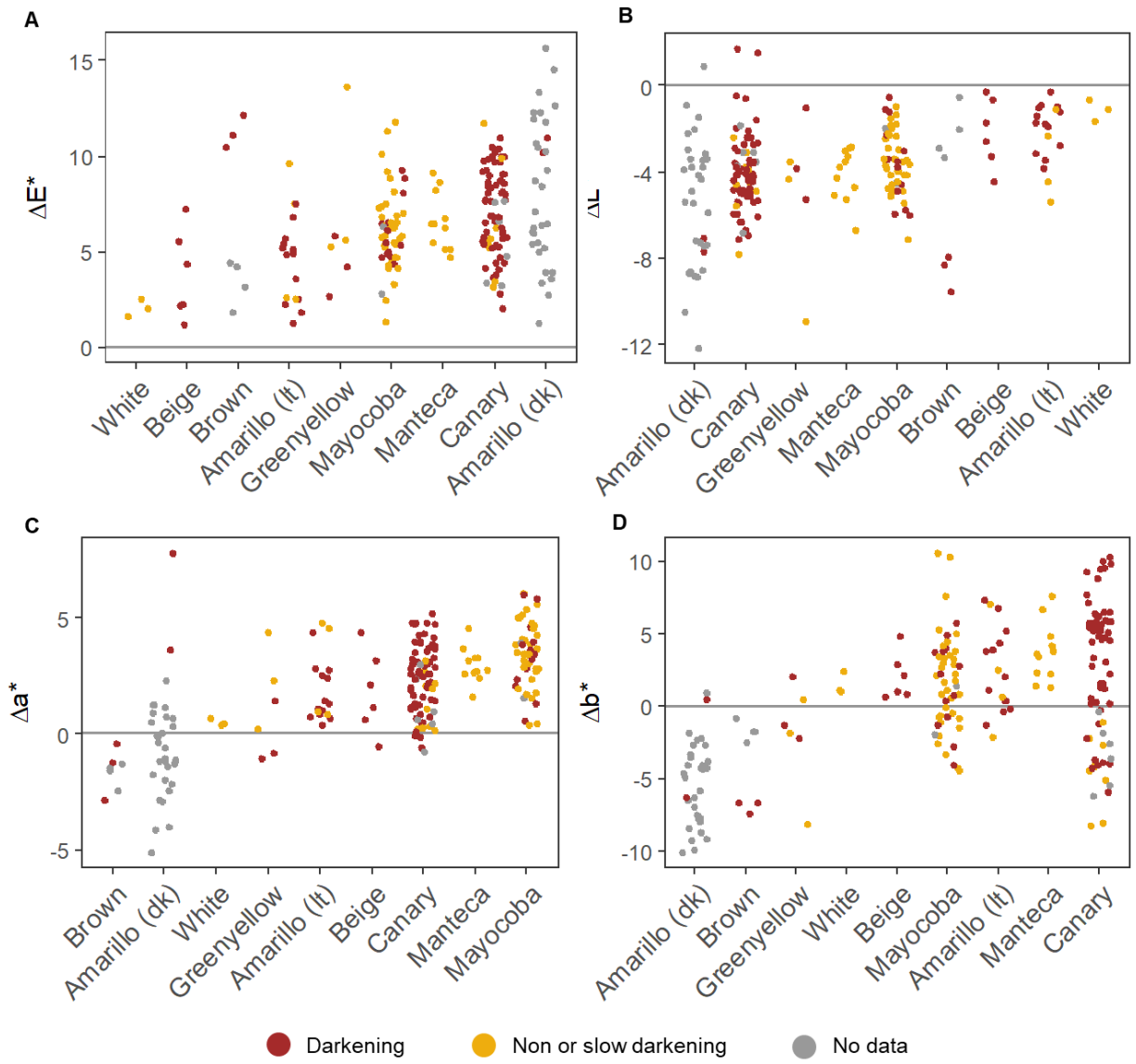


Figure S2.8 Color differences **A**: ΔE*, **B**: ΔL*, **C**: Δa*, and **D**: Δb* between MI and NE by seed type and postharvest darkening behavior.

REFERENCES

REFERENCES

Amatya, S., Karkee, M., Gongal, A., Zhang, Q., & Whiting, M. D. (2016). Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. *Biosystems Engineering*, *146*, 3–15. https://doi.org/10.1016/j.biosystemseng.2015.10.003

Arns, F. D., Ribeiro, N. D., Mezzomo, H. C., De Marco Steckling, S., Kläsener, G. R., & Casagrande, C. R. (2018). Combined selection in carioca beans for grain size, slow darkening and fast-cooking after storage times. *Euphytica*, *214*(4), 66. https://doi.org/10.1007/s10681-018-2149-8

Bassett, A., Katuuramu, D. N., Song, Q., & Cichy, K. (2021). QTL mapping of seed quality traits including cooking time, flavor, and texture in a yellow dry bean (*Phaseolus vulgaris* L.) population. *Frontiers in Plant Breeding*, (in revision).

Bassett, M. J., Shearon, C., & McClean, P. (1999). Allelism found between two common bean genes, hilum ring color (D) and partly colored seedcoat pattern (Z), formerly assumed to be independent. *Journal of the American Society for Horticultural Science*, *124*(6), 649–653. http://journal.ashspublications.org/content/124/6/649.short

Bassett, M J. (1999). The seedcoat color genotype of "Prim" and the Manteca and Coscorron market classes of common bean. *HortScience*, *34*(2), 336–337.

Bassett, M. J. (1996). The Margo (mar) Seedcoat Color Gene Is a Synonym for the Joker (j) Locus in Common Bean. *Journal of the American Society for Horticultural Science*, *121*(6), 1028–1031. http://journal.ashspublications.org/content/121/6/1028.abstract

Bassett, M. J. (2002). Classical and Molecular Genetic Studies of the Strong Greenish Yellow Seedcoat Color in `Wagenaar' and `Enola' Common Bean. *Journal of the American Society for Horticultural Science*, *127*(1), 50–55. http://journal.ashspublications.org/content/127/1/50.abstract

Bassett, M. J. (2007). Genetics of seed coat color and pattern in common bean. In *Plant Breeding Reviews* (pp. 239–315). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470168028.ch8

Bassett, M. J., & Miklas, P. N. (2007). A New Gene, *bic*, with Pleiotropic Effects (with *T P V*) for Bicolor Flowers and Dark Olive Brown Seed Coat in Common Bean. *Journal of the American Society for Horticultural Science*, *132*(3), 352–356. http://journal.ashspublications.org/content/132/3/352.full

Bassett, M. J., Miklas, P. N., Caldas, G. V., & Blair, M. W. (2010). A dominant gene for garnet brown seed coats at the Rk locus in 'Dorado' common bean and mapping Rk to linkage group 1. *Euphytica*, *176*(2), 281–290. https://doi.org/10.1007/s10681-010-0247-3

Bassett, M. J. (2003). Inheritance of yellow corona and hilum ring in seedcoats of Mayocoba market class common beans with genotype *P [C r] gy J g b* v^{lae} *Rk*. *Journal of the American*

Society for Horticultural Science, *128*(5), 721–723. https://doi.org/10.21273/JASHS.128.5.0721

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beebe, S. (2020). Biofortification of common bean for higher iron concentration. *Frontiers in Sustainable Food Systems*, *4*. https://doi.org/10.3389/fsufs.2020.573449

Beninger, C. W., Gu, L., Prior, R. L., Junk, D. C., Vandenberg, A., & Bett, K. E. (2005). Changes in polyphenols of the seed coat during the after-darkening process in pinto beans (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, *53*(20), 7777–7782. https://doi.org/10.1021/jf050051l

Beninger, C. W., & Hosfield, G. L. (1999). Flavonoid composition of three genotypes of dry bean (Phaseolus vulgaris) differing in seedcoat color. *Journal of Amer. Soc. Hort Sci*, *124*(5), 514–518.

Beninger, C. W., & Hosfield, G. L. (2003). Antioxidant activity of extracts, condensed tannin fractions, and pure flavonoids from *Phaseolus vulgaris* L. seed coat color genotypes. *Journal of Agricultural and Food Chemistry*, *51*(27), 7879–7883. https://doi.org/10.1021/jf0304324

Beninger, C. W., Hosfield, G. L., & Nair, M. G. (1998). Flavonol glycosides from the seed coat of a new Manteca-type dry bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, *46*(8), 2906–2910. https://doi.org/10.1021/jf9801522

Bennett, R. N., & Wallsgrove, R. M. (1994). Secondary metabolites in plant defence mechanisms. *New Phytologist*, *127*(4), 617–633. https://doi.org/10.1111/j.1469-8137.1994.tb02968.x

Bornowski, N., Song, Q., & Kelly, J. D. (2020). QTL mapping of post-processing color retention in two black bean populations. *Theoretical and Applied Genetics*, *133*(11), 3085–3100. https://doi.org/10.1007/s00122-020-03656-3

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bressani, R., & Elías, L. G. (1979). The nutritional role of polyphenols in beans. In J. H. Hulse (Ed.), *Polyphenols in cereals and legumes: 36th annual meeting of the Institute of Food Technologists* (pp. 61–68).

Castellanos, J. Z., Guzman-Maldonado, S. H., Jimenez, A., & Acosta-Gallegos, J. A. (1996). Preferential habits of common bean (*Phaseolus vulgaris* L.) consumers in Mexico. *Reports of Bean Improvement Cooperative and National Dry Bean Council Research Conference*, 182–183. https://naldc.nal.usda.gov/naldc/catalog.xhtml?id=IND20562914&start=0&searchText=IND20562914&searchField=&sortField=

Chen, P. X., Tang, Y., Marcone, M. F., Pauls, P. K., Zhang, B., Liu, R., & Tsao, R. (2015). Characterization of free, conjugated and bound phenolics and lipophilic antioxidants in regular- and non-darkening cranberry beans (*Phaseolus vulgaris* L.). *Food Chemistry*, *185*, 298–308. https://doi.org/10.1016/j.foodchem.2015.03.100

Cichy, K. A., Wiesinger, J. A., & Mendoza, F. A. (2015). Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *128*(8), 1555–1567. https://doi.org/10.1007/s00122-015-2531-z

CIE International Commission on Illumination. (2004). CIE 15: Technical Report: Colorimetry (3rd edition). https://www.cdvplus.cz/file/3-publikace-cie15-2004/

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *18*(3), 251–263. https://doi.org/10.1016/0169-7439(93)85002-X

Dixon, R. A., Xie, D.-Y., & Sharma, S. B. (2005). Proanthocyanidins – a final frontier in flavonoid research? *New Phytologist*, *165*(1), 9–28. https://doi.org/10.1111/j.1469-8137.2004.01217.x

Duitama, J., Quintero, J. C., Cruz, D. F., Quintero, C., Hubmann, G., Foulquié-Moreno, M. R., Verstrepen, K. J., Thevelein, J. M., & Tohme, J. (2014). An integrated framework for discovery and genotyping of genomic variants from high-throughput sequencing experiments. *Nucleic Acids Research*, *42*(6), e44. https://doi.org/10.1093/nar/gkt1381

Elia, F. M., Hosfield, G. L., Kelly, J. D., & Uebersax, M. (1997). Genetic analysis and interrelationships between traits for cooking time, water absortion, and protein and tannin content of andean dry beans. In *J. Amer. Soc. Hort. Sci.* (Vol. 122, Issue 4, pp. 512–518).

Elsadr, H. T., Wright, L. C., Peter Pauls, K., & Bett, K. E. (2011). Characterization of seed coat post harvest darkening in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *123*(8), 1467–1472. https://doi.org/10.1007/s00122-011-1683-8

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Engleright, R., Beimiriki, M., & Hosfield, G. (1999). Determination of total dietary fiber, indigestible starch, and indigestible protein in dry bean (*Phaseolus vulgaris* L.). *Annual Report of the Bean Improvement Cooperative*, *42*, 123–124.

Erfatpour, M., Navabi, A., & Pauls, K. P. (2018). Mapping the non-darkening trait from 'Wit-rood boontje' in bean (*Phaseolus vulgaris*). *Theoretical and Applied Genetics*, *131*(6), 1331–1343. https://doi.org/10.1007/s00122-018-3081-y

FAOSTAT. (2021). *No Title*. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#data/QC

Feenstra, W. J. (1960). *Biochemical aspects of seedcoat colour inheritance in* Phaseolus vulgaris

L. [Wargeningen University & Research]. https://library.wur.nl/WebQuery/wurpubs/525612

Fehr, W. R. (1987). *Principles of Cultivar Development*. Macmillan USA.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, *145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

González, A. M., Monteagudo, A. B., Casquero, P. A., De Ron, A. M., & Santalla, M. (2006). Genetic variation and environmental effects on agronomical and commercial quality traits in the main European market classes of dry bean. *Field Crops Research*, *95*(2–3), 336–347. https://doi.org/10.1016/j.fcr.2005.04.004

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, *127*, 418–424. https://doi.org/10.1016/j.compag.2016.07.003

Haeghen, Y. V., Naeyaert, J. M. A. D., Lemahieu, I., & Philips, W. (2000). An imaging system with calibrated color image acquisition for use in dermatology. *IEEE Transactions on Medical Imaging*, *19*(7), 722–730. https://doi.org/10.1109/42.875195

Hart, J. J., Tako, E., & Glahn, R. P. (2017). Characterization of polyphenol effects on inhibition and promotion of iron uptake by Caco-2 Cells. Journal of Agricultural and Food Chemistry, 65(16), 3285–3294. https://doi.org/10.1021/acs.jafc.6b05755

Hart, J. J., Tako, E., Wiesinger, J., & Glahn, R. P. (2020). Polyphenolic profiles of yellow bean seed coats and their relationship with iron bioavailability. *Journal of Agricultural and Food Chemistry*, *68*(3), 769–778. https://doi.org/10.1021/acs.jafc.9b05663

Hayat, I., Ahmad, A., Masud, T., Ahmed, A., & Bashir, S. (2014). Nutritional and health perspectives of beans (*Phaseolus vulgaris* L.): an overview. *Critical Reviews in Food Science and Nutrition*, *54*(February 2013), 580–592. https://doi.org/10.1080/10408398.2011.596639

Hooper, S., Wiesinger, J. A., Echeverria, D., Thompson, H. J., Brick, M. A., Nchimbi-Msolla, S., & Cichy, K. A. (2016). Carbohydrate profile of a dry bean (*Phaseolus vulgaris* L.) panel encompassing broad genetic variability for cooking time. *Cereal Chemistry Journal*, *94*(1), 135–141. https://doi.org/10.1094/CCHEM-04-16-0126-FI

Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, *8*(2). https://doi.org/10.1093/gigascience/giy154

Islam, F. M. A., Rengifo, J., Redden, R. J., Basford, K. E., & Beebe, S. E. (2003). Association between seed coat polyphenolics (tannins) and disease resistance in common bean. *Plant Foods for Human Nutrition*, *58*(4), 285–297. https://doi.org/10.1023/B:QUAL.0000040283.51023.c2

Islam, F. M. A., Beebe, S., Muñoz, M., Tohme, J., Redden, R. J., & Basford, K. E. (2004). Using

molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. *Theoretical and Applied Genetics*, *108*(2), 243–252. https://doi.org/10.1007/s00122-003-1437-3

Islam, N. S., Bett, K. E., Pauls, K. P., Marsolais, F., & Dhaubhadel, S. (2020). Postharvest seed coat darkening in pinto bean (*Phaseolus vulgaris*) is regulated by P^{sd}, an allele of the basic helix-loop-helix transcription factor *P*. *PLANTS, PEOPLE, PLANET*, *2*(6), 663–677. https://doi.org/10.1002/ppp3.10132

Junk-Knievel, D. C., Vandenberg, A., & Bett, K. E. (2007). An accelerated postharvest seed-coat darkening protocol for pinto beans grown across different environments. *Crop Science*, *47*(2), 694–700. https://doi.org/10.2135/cropsci2006.05.0325

Junk-Knievel, D. C., Vandenberg, A., & Bett, K. E. (2008). Slow darkening in pinto bean (*Phaseolus vulgaris* L.) seed coats is controlled by a single major gene. *Crop Science*, *48*(1), 189–193. https://doi.org/10.2135/cropsci2007.04.0227

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. https://cran.r-project.org/package=factoextra

Kelly, J. D., Awale, H. E., Wiersma, A. T., Cichy, K. A., & Wright, E. M. (2021). Registration of 'Yellowstone' yellow bean. *Journal of Plant Registrations*, 1–6. https://doi.org/10.1002/plr2.20075

Kelly, J. D., & Bornowski, N. (2018). Marker-Assisted breeding for economic traits in common bean. In S. S. Gosal & S. H. Wani (Eds.), *Biotechnologies of Crop Improvement, Volume 3* (pp. 211–238). Springer International Publishing. https://doi.org/10.1007/978-3-319-94746-4_10

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.5.5-1*. https://cran.r-project.org/package=emmeans

León, K., Mery, D., Pedreschi, F., & León, J. (2006). Color measurement in L∗a∗b∗ units from RGB digital images. *Food Research International*, *39*(10), 1084–1091. https://doi.org/10.1016/j.foodres.2006.03.006

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics*, *28*(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Lobaton, J. D., Miller, T., Gil, J., Ariza, D., de la Hoz, J. F., Soler, A., Beebe, S., Duitama, J., Gepts, P., & Raatz, B. (2018). Resequencing of Common Bean Identifies Regions of Inter–Gene Pool Introgression and Provides Comprehensive Resources for Molecular Breeding. The Plant Genome, 11(2). https://doi.org/10.3835/plantgenome2017.08.0068

Long, Y., Bassett, A., Cichy, K., Thompson, A., & Morris, D. (2019, June). Bean split ratio for dry bean canning quality and variety analysis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. https://openaccess.thecvf.com/content_CVPRW_2019/html/CVPPP/Long_Bean_Split_Ratio_for_Dry_Bean_Canning_Quality_and_Variety_CVPRW_2019_paper.html

Marles, M. A. S., Vandenberg, A., & Bett, K. E. (2008). Polyphenol oxidase activity and differential accumulation of polyphenolics in seed coats of pinto bean (*Phaseolus vulgaris* L.) characterize postharvest color changes. *Journal of Agricultural and Food Chemistry*, *56*(16), 7049–7056. https://doi.org/10.1021/jf8004367

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

MATLAB. (2020). *MATLAB: R2020a* (9.8.0.1538580). The MathWorks Inc.

McClean, P. E. (2002). Molecular and Phenotypic Mapping of Genes Controlling Seed Coat Pattern and Color in Common Bean (*Phaseolus vulgaris* L.). *Journal of Heredity*, *93*(2), 148–152. https://doi.org/10.1093/jhered/93.2.148

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

Mendoza, F. A., Kelly, J. D., & Cichy, K. A. (2017). Automated prediction of sensory scores for color and appearance in canned black beans (Phaseolus vulgaris L.) using machine vision. *International Journal of Food Properties*, *20*(1), 83–99. https://doi.org/10.1080/10942912.2015.1136939

Messina, V. (2014). Nutritional and health benefits of dried beans. *The American Journal of Clinical Nutrition*, *100*(1), 437S-442S. https://doi.org/10.3945/ajcn.113.071472.2

Mevik, B.-H., Wehrens, R., & Liland, K. H. (2020). *pls: Partial Least Squares and Principal Component Regression*. https://cran.r-project.org/package=pls

Osorno, J. M., Vander Wal, A. J., Kloberdanz, M., Pasche, J. S., Schroder, S., & Miklas, P. N. (2018). A new slow-darkening pinto bean with improved agronomic performance: registration of "ND-Palomino." *Journal of Plant Registrations*, *12*(1), 25–30. https://doi.org/10.3198/jpr2017.05.0026crc

Petry, N., Egli, I., Zeder, C., Walczyk, T., & Hurrell, R. (2010). Polyphenols and phytic acid contribute to the low iron bioavailability from common beans in young women. *The Journal of Nutrition*, *140*(11), 1977–1982. https://doi.org/10.3945/jn.110.125369

Possobom, M. T. D. F., Ribeiro, N. D., Zemolin, A. E. M., & Arns, F. D. (2015). Genetic control

of the seed coat colour of Middle American and Andean bean seeds. *Genetica*, *143*(1), 45–54. https://doi.org/10.1007/s10709-014-9811-4

Prakken, R. (1970). Inheritance of colour in Phaseolus vulgaris L. II. A critical review. *Mededelingen van de Landbouwhogeschool Te Wageningen*, *70*(23), 1–38.

Prakken, R. (1972). Inheritance of colours in *Phaseolus vulgaris* L. III. On genes for red seed coat colour and a general synthesis. *Mededelingen van de Landbouwhogeschool Te Wageningen*, *29*(1–82), 1–82.

Prakken, R. (1974). Inheritance of colours in *Phaseolus vulgaris* L. IV. Recombination within the "complex locus *C*." *Mededelingen van de Landbouwhogeschool Te Wageningen*, *24*, 1–36.

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Reddy, N. R., Pierson, M. D., Sathe, S. K., & Salunkhe, D. K. (1985). Dry bean tannins: A review of nutritional implications. *Journal of the American Oil Chemists Society*, *62*(3), 541–549. https://doi.org/10.1007/BF02542329

Ribeiro, N. D., Mezzomo, H. C., & Santos, G. G. dos. (2019). Genetic parameters and combined selection for seed coat color and macrominerals in Mesoamerican common bean lines. *Genetics and Molecular Research*, *18*(2). https://doi.org/10.4238/gmr18224

Rodríguez Madrera, R., Campa Negrillo, A., Suárez Valles, B., & Ferreira Fernández, J. J. (2020). Characterization of extractable phenolic profile of common bean seeds (*Phaseolus vulgaris* L.) in a Spanish diversity panel. *Food Research International*, *138*, 109713. https://doi.org/https://doi.org/10.1016/j.foodres.2020.109713

Sadohara, R., Izquierdo, P., Couto Alves, F., Porch, T. G., Beaver, J. S., Urrea, C. A., & Cichy, K. A. (2021). The *Phaseolus vulgaris* Yellow Bean Collection: Genetic diversity and characterization for cooking time (Under review). *Genetic Resources and Crop Evolution*.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. https://doi.org/10.1038/nmeth.2089

Silva, F. de C., Pereira, H. S., Melo, P. G. S., & Melo, L. C. (2018). Selection of parents and segregating populations of common bean with high agronomic potential and slow seed-coat darkening. *Pesquisa Agropecuária Tropical*, *48*(1), 75–82. https://doi.org/10.1590/1983-40632018v4849519

Takeoka, G. R., Dao, L. T., Full, G. H., Wong, R. Y., Harden, L. A., Edwards, R. H., & Berrios, J. D. J. (1997). Characterization of black bean (*Phaseolus vulgaris* L.) anthocyanins. *Journal of Agricultural and Food Chemistry*, *45*(9), 3395–3400. https://doi.org/10.1021/jf970264d

Tako, E., Beebe, S. E., Reed, S., Hart, J. J., & Glahn, R. P. (2014). Polyphenolic compounds appear to limit the nutritional benefit of biofortified higher iron black bean (*Phaseolus vulgaris* L.). *Nutrition Journal*, *13*(1), 28. https://doi.org/10.1186/1475-2891-13-28

Temple, S. R., & Morales, F. (1986). Linkage dominant hypersensitive resistance to bean common mosaic virus to seed color in *Phaseolus vulgaris* L. *Euphytica1*, *35*, 331–333.

Uebersax, M. A., & Siddiq, M. (2012). Market classes and physical and physiological characteristics of dry beans. In M. Siddiq & M. A. Uebersax (Eds.), *Dry Beans and Pulses Production, Processing and Nutrition* (pp. 55–74). Blackwell Publishing Ltd. https://doi.org/10.1002/9781118448298.ch3

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Varga, F., Vidak, M., Ivanović, K., Lazarević, B., Širić, I., Srečec, S., Šatović, Z., & Carović-Stanko, K. (2019). How does Computer vision compare to standard colorimeter in assessing the seed coat color of common bean (<i>Phaseolus vulgaris<i/> L.)? *Journal of Central European Agriculture*, *20*(4), 1169–1178. https://doi.org/10.5513/JCEA01/20.4.2509

Venables, W., & Ripley, B. (2002). *Modern applied statistics with S* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/

Voysest, O., Valencia, M. C., & Amezquita, M. C. (1994). Genetic diversity among Latin American Andean and Mesoamerican common bean cultivars. *Crop Science*, *34*, 1100–1110. https://doi.org/10.2135/cropsci1994.0011183X003400040049x

Voysest, Oswaldo. (2012). Yellow beans in Latin America. *Annual Report of the Bean Improvement Cooperative*, xii–xviii. http://bic.css.msu.edu/_pdf/Reports/BIC_2012_Annual_Report.pdf

Wiesinger, J. A., Glahn, R. P., Cichy, K. A., Kolba, N., Hart, J. J., & 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. https://doi.org/10.3390/nu11081768

Wiesinger, J. A., Osorno, J. M., McClean, P. E., Hart, J. J., & Glahn, R. P. (2021). Faster cooking times and improved iron bioavailability are associated with the down regulation of procyanidin synthesis in slow-darkening pinto beans (Phaseolus vulgaris L.). Journal of Functional Foods, 82, 104444. https://doi.org/10.1016/j.jff.2021.104444

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

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, *126*(2), 485–493. https://doi.org/10.1104/pp.126.2.485

Wortmann, C. S., Kirkby, R. A., Eledu, C. A., & Allen, D. J. (1998). Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa. *CIAT Publication*, *297*, 131.

http://hdl.handle.net/10568/54312

Wu, D., & Sun, D.-W. (2013). Colour measurements by computer vision for food quality control – A review. *Trends in Food Science & Technology*, *29*(1), 5–20. https://doi.org/10.1016/j.tifs.2012.08.004

Yang, Q.-Q., Gan, R.-Y., Ge, Y.-Y., Zhang, D., & Corke, H. (2018). Polyphenols in common beans (*Phaseolus vulgaris* L.): Chemistry, analysis, and factors affecting composition. *Comprehensive Reviews in Food Science and Food Safety*, *17*(6), 1518–1539. https://doi.org/10.1111/1541-4337.12391

Yin, L. (2020). *CMplot: Circle Manhattan Plot. R package version 3.6.2*. https://cran.r-project.org/package=CMplot

CHAPTER 3:

GENOTYPIC AND ENVIRONMENTAL EFFECTS ON PASTE QUALITY OF COMMON BEANS (PHASEOLUS VULGARIS L.) GROWN IN MICHIGAN

[Published in: HortScience 55(5): 684-692]

Genotypic and Environmental Effects on Paste Quality of Common Beans (*Phaseolus vulgaris* L.) Grown in Michigan

Rie Sadohara [1], James D. Kelly [1], Karen A. Cichy [1,2]

1. The Department of Plant, Soil and Microbial Sciences, Michigan State University, 1066 Bogue St. East Lansing, MI. 48824, USA
2. USDA-ARS, Sugarbeet and Bean Research Unit, 1066 Bogue St. East Lansing, MI. 48824, USA

Abstract

Common beans are recognized as a nutrient dense food source that delivers numerous health benefits, but one of the barriers to increasing bean consumption is the limited number of common bean food products. Bean paste, made from bean seed and sugar, has the potential to diversify and expand the way beans are consumed. In this study, commercial white seeded otebo, navy, great northern, and white kidney bean cultivars, and one colored cranberry bean were grown in two environments in Michigan and evaluated for bean paste qualities. Characteristics such as paste yield, color, flavor, and stickiness were evaluated on the bean paste. The genotype × environment effect was significant for many of the paste making qualities and the color values of the unsweetened paste. Snowdon, the white kidney bean, had superior paste yield of unsweetened paste and whiteness of sweetened paste in both environments. All the white bean cultivars were comparable to Hime, the control otebo cultivar, in terms of low flavor intensity. Powderhorn, the great northern bean, had high stickiness of sweetened paste, which is preferable. The cranberry bean resulted in dark-colored paste with high flavor intensity. Seed coat percentage and the ratio of L* and C* obtained via image analysis could be used as indicators for paste yield and whiteness score of the unsweetened paste, respectively. Overall, these results suggested that specific

domestically grown white bean cultivars have potential for development as bean paste products, which would add a novelty to the processed dry bean applications in the U.S.

Introduction

Common beans (*Phaseolus vulgaris* L.) are a good source of protein, dietary fiber, vitamins, and minerals and offer numerous health benefits to consumers (Hayat et al., 2014; Messina, 2014). However, the annual per capita consumption of beans was only 5.8 pounds (2.6 kg) during 2011-2015 (US Department of Agriculture Economic Research Service, 2017). One of the reasons for the low bean consumption is the limited number of product applications for which beans can be used (Desrochers & Brauer, 2001; Smith et al., 2016). Originating in Asia, bean paste is a confectionery ingredient made of beans and sugar. Otebo beans are a market class that is developed specifically for paste-making purposes, but other white common bean classes such as great northern can also be used (Kamiya et al., 2004). Bean paste can be a novel use of dry beans in the U.S., which may promote bean consumption especially because of today's increasing demand for pulse-based products (Tyler et al., 2017). Bean paste can be prepared with or without the seed coat; the latter results in paste made solely from cooked cotyledon with a smooth texture (Sadohara, 2019). Otebo (a.k.a. Tebou) beans are mid-sized white bean cultivars used for white paste production in Japan (Kato, 2000), and bean paste qualities are extensively studied for breeding this market class (Kato, 2000; Komiyama, 2013; Komiyama & Kato, 2004).

Important bean paste qualities include paste yield, whiteness, stickiness, smoothness, and flavor described in detail by Sadohara (2019). Paste yield is considered when paste is made without the seed coat and is an indicator of the efficiency of paste production. High whiteness, stickiness, and smoothness are preferred for smooth white paste. For paste flavor, lack of strong beaniness

(bean-like flavor) or grassiness (green, grass-like flavor) is considered desirable. Paste yield is reported to be correlated with some seed and cooking quality traits (Kato, 2000), whereas other paste quality traits can only be measured by paste preparation and evaluation by a sensory panel. Because of the labor-intensive nature of the evaluation process, predictive and high-throughput methods have been explored to accelerate the breeding process (Komiyama, 2013; Komiyama & Kato, 2004).

Several otebo market class cultivars were developed in Michigan to be exported for bean paste production (Kelly et al., 2009, 2016). For example, otebo bean cultivar Samurai was developed with taller upright architecture to suit direct machine harvest (Kelly et al., 2016). However, the potential of otebo beans and other white market classes for bean paste has not been evaluated due to the lack of screening tools to select for paste traits. Genotype × environment effects have been reported for bean paste quality traits (Komiyama and Kato, 2004), so these effects are also important considerations for bean growers across the state and the nation. The current study aimed to evaluate existing white bean cultivars for paste qualities, to investigate the effects of growing environments on the quality parameters, and to examine phenotypes useful for predicting paste qualities without the need for paste preparation.

Materials and methods

Plant materials

Genotypes tested comprised five commercial white bean cultivars: Hime (otebo) (Miura et al., 1977), Samurai (otebo) (Kelly et al., 2016), Alpena (navy) (Kelly et al., 2015), Powderhorn (great northern) (Kelly et al., 2014), and Snowdon (white kidney) (Kelly et al., 2012), and CR1502-4 (a cranberry bean) (**Table 3.1**). Hime was used as a control cultivar because it is an otebo bean

cultivar used for commercial bean paste production (Kelly et al., 2009). The cranberry line was included because of its tendency to disintegrate during canning (unpublished data). The ability to break easily upon cooking is expected to be a suitable characteristic for paste making, which involves cooking beans until they start to disintegrate. The beans were grown as described by Kelly & Cichy (2012) at two experimental stations, Montcalm (MC, 43°21'09.5"N, 85°10'36.2"W) and Saginaw Valley (SV, 43°24'12.6"N, 83°41'58.8"W) in Michigan in 2017. The experiment was planted in a randomized complete block design with two field replications. At MC, each plot (replication) had two 6.1 m-rows with 0.51 m row width, and genotypes were sown at the rate of 80 seeds/row. The 4.6 m-section of the two rows were harvested at maturity. At SV, each plot had four 6.1 m-rows with 0.51 m row width. Genotypes were sown at the rate of 80 seeds/row in the center 2 rows, and a dark red kidney bean was sown in the outer two rows serving as borders. The 4.6 m-section of the center two rows were harvested at maturity. Fertilizers were applied according to the local practices; 15-5-13 + S, Zn, Mn, Cu prior planting at SV and 19-10-19 prior to planting and 46-0-0 on day 27 at MC. Due to a low seed yield at SV, Hime was planted in two plots in each block, and the seeds from the two plots were combined. Throughout the growth period, a total of 137 mm of overhead irrigation was applied to supplement rainfall at MC, but not at SV. Beans were harvested at their full maturity with a Hege plot thresher. After harvest, beans were dried at room temperature, cleaned to remove chaff and large stones by using a seed cleaner (Clipper Office Tester, AT Ferrell Company Inc., Bluffton IN, USA), hand-cleaned to remove small stones and moldy or foreign seeds, and stored at a low temperature and low humidity chamber until the experiment was carried out. Before cooking, seed moisture was equilibrated in a cold storage room at 4°C and 75% relative humidity and confirmed to be in the range of 10-14% before the samples were subjected to paste preparation.

Seed quality

The hundred-seed weight of each cultivar was determined as an average of three measurements for each field replication. Seed moisture was determined by drying 10 seeds in a convection oven at 105°C for 72 h with two replications. Seed coat percentage was determined by soaking 15 seeds in distilled water for 16 h, separating the coat from the cotyledon, and drying them at 105C° for 24 h. The coat percentage was determined as (coat weight, dry weight basis [dwb]) / (coat plus cotyledon weight, dwb). Seed moisture and coat percentage of each plot was determined as an average of two measurements. Soaked seed hardness was determined by using a TA.XTplus100 with Exponent software ver. 6,1,11,0 (Stable Micro Systems Ltd, Godalming, UK). A sample of 20 seeds per field replication were soaked in distilled water for 16 h, and each seed was punctured by a cylindrical ϕ2 mm-probe, TA-52, with a compression rate of 70% with the pre-test, test, and post-test speeds of 1, 2, and 2 mm/s, respectively. Peak positive force and peak positive area were recorded for each sample.

Bean paste preparation and cooking quality evaluation

For each field replication, paste was prepared with two technical replications except CR1502-4 grown at SV, from which one paste sample was lost. Bean paste was prepared according to the method of Kato (2000) with modifications. Namely, 70 g of raw beans were soaked in distilled water for 16 h and drained to determine water uptake, calculated as (absorbed water weight) / (raw bean weight) ×100. Soaked beans were cooked in 2.5 L of distilled boiling water for 45 min. The cooked beans were drained and weighed to determine the Weight Increase Rate by Boiling (WIRB), which was calculated as (boiled bean weight) / (raw bean weight, dwb). The drained cooked beans were then mashed with a wooden spatula for 8 min and passed through a 0.5 mm-sieve to remove

the seed coat. The cotyledon was washed with 800 mL distilled water, rested for 30 min, and the supernatant was discarded. Washing with 500 mL distilled water followed by 10 min rest was repeated twice. The sediment was squeezed dry in a 50 cm × 50 cm cheesecloth to remove excess moisture. The resultant unsweetened paste was weighed, and the moisture content was determined by drying 1 g of paste in a convection oven at 105C° for 24 h in two replications. Paste yield was calculated as (unsweetened paste weight, dwb) / (raw bean weight, dwb). Unsweetened paste was sealed in plastic wrap and rested at 4C° for 14 h before it was sweetened. Sweetened paste was prepared under a consistent heating condition on an electric burner (CB-30P1,Cuisinart, Stamford, CT) with a 45 g of unsweetened paste with 45 mL of distilled water and 27.5 g of granulated sugar until the paste was ×1.3 of the unsweetened paste weight (58.5±0.1 g). Sweetened paste was immediately sealed in plastic wrap and cooled to room temperature. The moisture content of sweetened paste was determined using the same methodology as unsweetened paste by using 2 g of sweetened paste samples. The color, stickiness, and whiteness were evaluated on the sweetened paste within 10 h of preparation.

Color

The color of unsweetened and sweetened paste was measured by taking a digital image of 42 g paste filled in a ϕ60 mm-glass petri dish and flattened by a spatula under the imaging conditions described by Mendoza et al. (2017) except that the shutter speed was set to 1/100 s. Image J software (Abramoff et al., 2004) was used to extract L*, a*, and b* values (CIE International Commission on Illumination, 1978), from a ϕ56.4 mm-circle specified as a region of interest of each image. L* is a luminance component value, ranging from complete black of 0 to complete white of 100. Negative and positive values of a* and b* represent green-red and blue-yellow,

respectively (Yam & Papadakis, 2004). Chroma (C*) was determined as $C* = \sqrt{a*^2 + b*^2}$. The ratio L*/C* was calculated to assess the lightness of a paste sample in relation to its redness and yellowness.

Sensory evaluations

Sweetened and unsweetened samples were shaped into a 20 mm x 30 mm rectangular shape with 4 mm-thickness and were subjected to sensory evaluations. Four trained assessors evaluated eight samples per session. The assessors evaluated the whiteness, beaniness, vegetativeness, sweetness, and the total flavor intensity of unsweetened paste and the whiteness of sweetened paste on a 1-5 scale with 5 being the strongest. The samples' identity was not disclosed during the evaluation. The protocol was approved by the Michigan State University Biomedical, Health Sciences Intuitional Review Board (IRB): IRB# x16-763e Category: Exempt 6.

Stickiness

Stickiness was measured according to the method of Komiyama & Kato (2004) with slight modifications. Sweetened paste (42 g) in a ϕ60 mm-petri dish described above was compressed by a texture analyzer, TA.XTplus100 with Exponent software ver. 6,1,11,0 (Stable Micro Systems Ltd, Godalming, UK), equipped with a platform (TA-90) and a ϕ25.4 mm-cylindrical probe (TA-11) at a compression rate of 70% and with the pre-test, test, and post-test speeds of 2, 2, and 1 mm/s, respectively. Stickiness was recorded as the negative area (g·sec) where the probe is pulled upwards after compression.

Data analyses

The basic statistics of each trait were calculated by using the FSA package (Ogle et al., 2019) in the R platform (R Core Team, 2017). The scattered plot and the correlation coefficient of related traits were generated using the PerformanceAnalytics package (Peterson & Carl, 2019) in the R platform.

The PROC MIXED procedure of SAS software (ver. 9.4, SAS institute, USA) was used to estimate the genotype, environment, and genotype × environment interaction effects and to compare the means of each sample within the environments. To analyze all the quality traits except the sensory attributes across the environments, a model of $Y_{ijk} = \mu + G_i + E_j + GE_{ij} + B(E)_{jk} + \varepsilon_{ijk}$ was used, where: Y_{ijk} is the trait value of the *i*th genotype grown in the *k*th block of the *j*th environment, μ is the grand mean, G_i is the fixed effect of the *i*th genotype, E_j is the fixed effect of the *j*th environment, GE_{ij} is the interaction term of the *i*th genotype and the *j*th environment, $B(E)_{jk}$ is the fixed effect of the *k*th block nested in the *j*th environment, and ε_{ijk} is the error term. To analyze the sensory attributes across the environments, a random effect of *l*th subsamples (S_l), which was nested in block, was added to the model: $Y_{ijkl} = \mu + G_i + E_j + GE_{ij} + B(E)_{jk} + S_{l(jk)} + \varepsilon_{ijkl}$. To estimate and compare the means of genotypes within each environment, a model of $Y_{ik} = \mu + G_i + B_k + \varepsilon_{ik}$ was used for all the traits except the sensory attributes, for which $Y_{ikl} = \mu + G_i + B_k + S_{l(k)} + \varepsilon_{ikl}$ was used. Normality of the residuals was evaluated by visual inspection of the quantile-quantile plots and histograms. The PROC GLIMMIX procedure was used to group the genotypic means within an environment. The PROC CORR procedure was used to estimate the correlation coefficient between related traits measured across the environments. All tests were carried out at α=0.05 and mean comparison was done with Tukey adjustment.

Results

Seed quality

A graphic of the unprocessed dry seeds of each cultivar harvested from the two environments, MC and SV, is shown in **Table 3.1**. The seeds of Hime, Snowdon, and CR1502-4 from SV were less plump and the seed surface was uneven compared to the seeds from MC. These differences suggest that the growing conditions were less favorable at SV than at MC for certain genotypes in 2017.

Table 3.2 shows the genotype (G), environment (E), and genotype × environment (G×E) effects of the seed and paste qualities of each genotype grown in the two environments, MC and SV. G, E, and G×E effects were significant for all the seed qualities. **Table 3.3** shows the means of the seed and paste qualities of each genotype grown in the two environments. All the genotypes had the same 100 seed-weights in both environments except Snowdon and CR1502-4, which had lower seed weights at SV. Genotypes grown at MC had lower seed coat percentages than those grown at SV in general. There was a G×E effect on the seed coat percentage (**Table 3.2**); Alpena, Snowdon, and CR1502-4 had lower seed coat percentages than Hime at MC, whereas Samurai and Alpena were lower than Hime at SV (**Table 3.3**). For both the peak positive force and the positive area, MC samples had higher means than SV samples (**Table 3.3**). In addition, MC samples had a lower coefficient of variation than SV samples for the hardness measurements of soaked seeds (**Table 3.2**). Snowdon and CR1502-4 at MC had low peak positive force of soaked seeds among the genotypes, but Snowdon and Hime had low peak positive force at SV (**Table 3.3**). The positive area measured the total force required to compress the soaked beans. Hime had the highest positive area at MC, but it was one of the lowest at SV. Alpena and Snowdon had lower values than other genotypes in both environments.

Cooking and paste-making qualities

G, E, and G×E effects were all significant for the paste-making qualities (**Table 3.2**). Powderhorn and Snowdon had a similar water uptake to Hime in both environments. CR1502-4 had a comparable water uptake to Hime at SV (**Table 3.3**). The two otebo cultivars, Hime and Samurai, had the highest WIRB at MC and SV, respectively. Differences in water uptake and WIRB among the samples were more pronounced at SV than MC (**Table 3.2**). Snowdon had the highest paste yield, but other genotypes including Samurai had a comparable paste yield to Hime at MC. The five white-seeded cultivars had paste yields equivalent to that of Hime, whereas CR1502-4 was different from the other five cultivars at SV in terms of paste yield (**Table 3.3**). **Figure 3.1** shows the correlation between the paste yield and the seed and cooking qualities. Paste yield was negatively correlated with the seed coat percentage and positively correlated with soaked hardness (peak positive force and positive area) (**Figure 3.1**). WIRB was not strongly correlated with paste yield (**Figure 3.1**, r=0.35).

Whiteness and the color values of unsweetened paste

All the color values of unsweetened paste samples had G, E, and G×E effects except the E effect for the a* value and the interaction effect for the L* and a* values (**Table 3.2**). The unsweetened paste made from all the genotypes had a similar L* value except CR1502-4 in both environments (**Table 3.3**). Snowdon and Powderhorn had lower a* values while CR1502-4 had a high a* value in both environments. Alpena, Samurai, and CR1502-4 had low b* and C* values that were comparable to Hime, while Powderhorn had high b* and C* values at MC. There was no difference among genotypes grown at SV for b* and C* values. Samurai and Alpena were the two highest for L*/C* in both environments.

A sensory panel evaluated paste whiteness on a 1-5 scale with 5 being the whitest. There were G, E, and G×E effects for the whiteness of unsweetened paste (**Table 3.2**). Samurai, Alpena, and Snowdon had similar whiteness to Hime, and Powderhorn had a lower score at MC, whereas; at SV, Powderhorn was the whitest, and Hime and Samurai had relatively low whiteness scores (**Table 3.3**). In both environments, unsweetened paste made from CR1502-4 was by far the darkest.

Figure 3.2A shows the correlation coefficients and the scatter plots between the whiteness score and the color values obtained by the image analysis of unsweetened paste. When all the 47 paste samples were used, the whiteness of unsweetened paste was strongly correlated with L* and a* values, but it was due to CR1502-4 which had a darker paste with low L* and high a* values, as compared to the other genotypes. Therefore, the correlation between the whiteness score and the color values were examined after the seven CR1502-4 paste samples were removed (**Figure 3.2B**, n=40). The whiteness score had a positive correlation with L* and L*/C* and a negative correlation with a*, b*, and C* as expected. The L*/C* had a higher correlation coefficient than L* with the whiteness score.

Whiteness and the color values of sweetened paste

The genotype effects were significant for all the color values of sweetened paste, but the environmental effects were significant only for a*, b*, and C* values (**Table 3.2**). G×E effects were significant only for the a* value. Similar to the unsweetened paste, all white-seeded genotypes produced sweetened paste with similar L* values, while CR1502-4 had a low L* value. Hime, Samurai, and Alpena had similar a*, b*, C*, and L*/C* values (**Table 3.3**). Snowdon had low mean values of a*, and Powderhorn had high b* values in both environments. Snowdon had the highest L*/C* value at MC, while L*/C* was not different among the genotypes at SV.

The whiteness scores evaluated by assessors had G and G×E effects (**Table 3.2**). All the sweetened paste made from the white-seeded genotypes had a similar score as Hime, but Snowdon was the whitest, producing its highest L*/C* value at MC (**Table 3.3**). Powderhorn and Snowdon had higher whiteness scores than Hime, Samurai, and Alpena at SV. CR1502-4 was the darkest for the whiteness evaluation in both environments as was also observed with the unsweetened paste evaluation.

Figure 3.3A shows the correlation coefficients and the scatter plots between the whiteness score and the color values obtained by the image analysis of sweetened paste made from all the six genotypes (n=47). Similar to unsweetened paste, the L*, a*, b*, and C* values were correlated with the whiteness score, but it was due to the paste made from CR1502-4 that had lower L* and b* values than other genotypes; therefore, the data points of CR1502-4 from both environments were removed (**Figure 3.3A**, 3B, n=40). When CR1502-4 was removed, the whiteness score was not strongly correlated with any color values or the composite values (**Figure 3.3B**).

Flavor of unsweetened paste

The beaniness, vegetativeness, sweetness, and the total flavor intensity of the unsweetened paste were evaluated on a 1-5 scale. The coefficient of variation was high for these flavor attributes in both environments compared to other traits (**Table 3.2**). There was G effect on the beaniness and the total flavor intensity, and G×E effects on the total flavor intensity. There were no G, E, or G×E effects on vegetativeness and sweetness. The genotypes were different in beaniness and in the total flavor intensity only at MC (**Table 3.3**). Samurai, Alpena, and Powderhorn had similar beaniness and total flavor intensity scores as Hime, whereas Snowdon and CR1502-4 had greater scores than

other cultivars at MC. All the genotypes had similar and low scores for sweetness in both environments.

Stickiness of sweetened paste

There was a genotypic effect on the stickiness of sweetened paste, but the environment and the interaction effects were not significant (**Table 3.2**). All samples at MC had similar values to Hime; and Alpena, Snowdon, and Powderhorn had a comparable stickiness to Hime at SV. Powderhorn had the highest stickiness in both environments. The genotypes could be divided into two groups based on their stickiness values: Hime, Powderhorn, and Snowdon with high stickiness and Samurai, Alpena, and CR1502-4 with low stickiness.

Discussion

Cooking- and paste-making qualities in relation to seed quality

Bean paste has the potential to be a novel end-use of dry beans, which would lead to increased bean consumption. Therefore, this experiment tested the bean paste qualities of existing cultivars and a breeding line developed in Michigan in comparison to Hime, the industry standard developed in Japan. The cultivars were grown in two contrasting environments in Michigan, one irrigated (MC), the other rainfed (SV).

Paste yield, the efficiency of producing unsweetened paste from raw seeds, is an important characteristic for bean paste producers. Snowdon had high paste yield at MC and a comparable yield to Hime at SV. Based on these two-environment data, the suitability of Snowdon for paste production is promising. The paste yield of CR1502-4 was environment dependent; it had high paste yield at MC and low paste yield at SV. Other genotypes had a comparable paste yield to

Hime in both environments. It is noteworthy that the coefficient of variation for paste yield was higher at SV than MC, suggesting that some genotypes produce less uniform paste when grown at SV. In addition, all the genotypes had less paste yield at SV than MC. The irrigation at MC may have produced seeds that are easier to cook and to separate the seed coat during paste preparation. Despite the low paste yield, some genotypes from SV had a higher water uptake presumably because some of the genotypes from this environment had cracked and/or peeled seeds during soaking, which made further imbibition easier. However, the high water uptake of SV seeds did not lead to high paste yield; on the contrary, as the low WIRB indicated, the seeds were less cooked resulting in less separation of cotyledon cells and low paste yield. Samurai was an exception as it had a WIRB as high as Hime at SV (**Table 3.3**). Samurai was bred as an otebo bean, and its seed characteristics derived from the otebo pedigree may have resulted in high water uptake during cooking, a characteristic preferable for bean paste production. A related study of a black and pinto bean indicated that the starch granules of the two cultivars grown under rainfed conditions had a high ability to absorb water and swell but also were more susceptible to shear and granule rupture than those grown under irrigation (Ovando-Martínez et al., 2011). Although the current study did not include black or pinto beans, the environment and water regime likely influenced the cooking quality of the beans tested.

A previous study reported a positive correlation between WIRB and paste yield (Kato, 2000). The low correlation between WIRB and paste yield in this study (**Figure 3.1**) was deemed to be due to one of the CR1502-4 paste samples grown at SV that had an extremely low paste yield (28.1). This was a moderate outlier (>1.5 times of the interquartile range for this environment), and the correlation was higher (r=0.52, p=0.010) when this data point was removed. Thus, WIRB could be a useful indicator of paste yield when genotypes that had extremely low paste yield are

excluded from the analysis. The seed coat percentage was negatively correlated with paste yield, as expected, and could be another indicator of paste yield (**Figure 3.1**). From a practical point of view, the seed coat percentage would be a better indicator of paste yield than WIRB as it is easier to measure and does not involve the labor-intensive cooking process. However, very small beans such as navy may be outside the range of prediction because Alpena did not have a high paste yield at MC despite its low seed coat percentage (**Table 3.3**). Navy beans in the U.S. have long been bred for high tolerance to seed coat splitting during canning operation, and the seed coat of the tolerant navy lines tightly adhere to the cotyledon (Dorrell & Adams, 1969). This may explain why the seed coat removal after cooking of Alpena was not as efficient during the paste preparation.

Whiteness and color values of unsweetened and sweetened paste

The L* (lightness), a* (redness), and b* (yellowness) values indicated that the unsweetened paste made from Hime, Samurai, and Alpena tend to be redder while the unsweetened paste made from Powderhorn and Snowdon tend to be more yellow than red (**Table 3.3**). This trend was mostly the same with the sweetened paste except with Powderhorn grown at MC, which had the highest a* value. CR1502-4 resulted in very dark-colored paste unlike the other white genotypes. Cranberry beans have red pigments similar to Adzuki beans used for red paste; however, the color values of unsweetened paste of CR1502-4 were different from those typically found with Adzuki unsweetened paste (L*: 40, a*: 6, b*: 6) (Tazawa et al., 2015). There is no precedence of making paste from cranberry beans, so this result has provided new information on the end-use characteristic of this market class.

The whiteness scores of sweetened paste made from the white-seeded bean cultivars indicated that all those cultivars are likely to produce similar sweetened paste to Hime (**Table 3.3**).

The high whiteness score of sweetened paste made from Snowdon in both environments indicated that this cultivar may withstand darkening during the sweetening process. High whiteness of sweetened paste is highly valued because it directly impacts the confectionery product quality for which bean paste is used. Hime and Powderhorn produced paste with different whiteness scores depending on the environment, indicating that the growing environment plays an important role on paste color for some genotypes.

Prediction of whiteness using color values

Considering that all the unsweetened paste had similar L* values except CR1502-4, the L* value alone is not sufficient to capture the overall impression on the unsweetened paste color evaluated by the sensory panel (**Table 3.3**). Therefore, composite values such as C* and L*/C* were examined. C* and L*/C* were influenced largely by b* because the paste samples had larger b* values than a* (**Table 3.2**). The L*/C* values could be used to predict whiteness scores given by a sensory panel on unsweetened paste made from white beans. Color values such as L*, a*, and b*, measured using colorimeters, have been used for evaluating bean paste color (Okuyama et al., 2008; Shinada et al., 1994), but no study has examined machine vision and image analysis technology for this purpose. Computer vision is widely used for evaluating the color of food products at a high resolution (Wu & Sun, 2013). This method could be used for an objective and efficient screening of breeding lines for white paste color without the need for training assessors.

The L*/C* values of sweetened paste differentiated genotypes grown at MC but not at SV despite the difference between them detected by the assessors (**Table 3.3**). Still, Snowdon had a high L*/C* and high whiteness score at MC; thus, L*/C* may be useful in identifying outstandingly white sweetened paste. L*/C* or other color values were not strongly correlated with

the whiteness score of the five white genotypes (**Figure 3.3B**). One possible reason for this is that L*, b*, C*, and L*/C* of sweetened paste had higher coefficients of variation than those of unsweetened paste (**Table 3.2**), indicating that within-sample variability increased when paste was sweetened. This is probably because the sweetening process involves manual stirring (Shiota & Miyata, 1976) and is difficult to perform consistently. In order to rank genotypes according to the whiteness score of sweetened paste, a trained panel of assessors may still be necessary, but it is a costly way to evaluate a large number of samples. The whiteness score of unsweetened paste was positively correlated with that of sweetened paste made from the white genotypes (r=0.45, p=0.0033); thus, an economical approach might be to evaluate unsweetened paste by comparing the L*/C* values first obtained by image analysis thereby indirectly selecting for high whiteness of sweetened paste. Later in the selection, a sensory panel could evaluate sweetened paste samples made from a smaller number of elite lines.

Flavor of unsweetened paste

None of the flavor attributes had significant E (environment) effects, indicating that the flavor attributes are not influenced by environment. Nonsignificant environmental effects on the flavor were consistent with a previous study that compared the aroma and flavor of six solid red, red- or dark-striped beans (Mkanda et al., 2007). The high CV of the flavor attributes was due to the high variability among the assessors. Unlike technical replications of other trait measurements, higher variability in sensory evaluation is expected because the assessors are individuals with varying sensitivity to different flavor attributes. However, the way assessors differentiated the samples aligned with the expected results such as the low flavor intensity of Hime, the control cultivar, and the high flavor intensity of non-white, cranberry genotype at MC (**Table 3.3**). The similar and low

flavor intensity among the white beans suggested that any white-seeded genotypes are comparable to Hime in terms of flavor despite Snowdon being higher in beaniness at MC.

The unsweetened paste made from CR1502-4 had the highest beany flavor compared to other white beans at MC. This finding was in line with a previous study where two maroon-mottled and two solid beige beans from Europe tended to have stronger flavor than solid white beans when tasted as whole boiled beans (Rivera et al., 2013). The authors also reported that the flavor intensity was negatively correlated with the percentage of white color on the seeds in their collection, which consisted of 20 accessions of various colors. It may explain the stronger flavor detected in the cranberry genotype in this study on the assumption that the flavor characteristics of whole beans reflect those of unsweetened paste. All genotypes had similar and low vegetativeness and sweetness in both environments, indicating that those attributes of bean paste are too subtle and indistinguishable between the genotypes. The similar scores of sensory attributes among the paste samples at SV may indicate that beans would produce paste with more similar flavor grown in this environment. However, considering the high coefficient of variation of the soaked hardness and paste making-qualities of the genotypes at SV, it is also possible that the paste samples had high within-genotype variability that led to reduced statistical power to detect differences between the genotypes. Further research will be necessary to test this hypothesis.

Stickiness of sweetened paste

The significant G effect and the insignificant E and G×E effects indicated that genotype is an important factor for this trait. The high stickiness of Powderhorn in both environments is a favorable characteristic for bean paste made without the seed coat. Stickiness is evaluated by preparing paste, and no phenotype was suggested as a predictor of stickiness. However, the protein

content of sweetened paste was correlated with stickiness for some otebo cultivars and breeding lines (Komiyama & Kato, 2004). As a future study, the protein content of sweetened and unsweetened paste can be measured to facilitate the understanding of the genotypic differences in paste stickiness. The moisture contents of the sweetened paste samples were not correlated with stickiness (p=0.75); thus, it did not explain the genotypic differences.

Conclusions

Bean paste could be a novel use of dry beans, and this study revealed the potential of Michigan-grown, white-seeded beans for paste production. All the white cultivars tested performed similar to or better than Hime for many of the paste quality traits such as paste yield, L* and L*/C* values of the unsweetened paste, and L* values of the sweetened paste. The large-seeded white beans such as Powderhorn and Snowdon showed the best potential for some of the paste traits. These commercial cultivars are currently used for other purposes such as canning, but they could also perform well as paste; therefore, if there is increased demand for bean paste in the U.S. or other parts of the world, it can be met with existing cultivars. Growing environments and/or the use of irrigation influenced the seed and paste qualities, so this should be an important consideration. With the promising genotypes and evaluation methods identified in this study, future experiments can be designed to evaluate and validate the performance of future bean cultivars and their suitability for paste production.

Acknowledgements

We thank Filipe Couto Alves at CANR Biometry Group Statistical Consulting Center, Michigan State University for advice on the statistical analyses.

APPENDICES

APPENDIX A:

CHAPTER 3 TABLES AND FIGURES

Table 3.1 Six common bean genotypes tested for agronomic and paste quality traits in two environments in Michigan.

Genotype	Hime	Samurai	Alpena	Powderhorn	Snowdon	CR1502-4
Market class	Otebo	Otebo	Navy	Great northern	White kidney	Cranberry
Seed color	White	White	White	White	White	Cream with red mottling
Type	Cultivar	Cultivar	Cultivar	Cultivar	Cultivar	Breeding line
Unprocessed seeds (Montcalm)						
Unprocessed seeds (Saginaw Valley)						

Table 3.2 The genotype (G), environment (E), and genotype × environment (G×E) effects on the traits measured and the basic statistics of each trait in two environments.

	(unit)	G	E	G x E	Environment	n	Min.	Median	Max.	Mean	S.D.[z]	C.V.[y] (%)
Seed quality												
100 seed weight	(g)	<0.0001	<0.0001	<0.0001	MC[x]	24	16.8	30.1	66.2	38.0	19.3	50.6
					SV[w]	24	17.0	27.6	52.9	33.0	12.7	38.5
Seed coat	(%)	<0.0001	<0.0001	<0.0001	MC	24	6.3	7.3	8.0	7.2	0.4	5.6
					SV	24	6.8	7.8	9.5	8.0	0.7	8.9
Peak Positive Force[v]	(g)	<0.0001	<0.0001	<0.0001	MC	240	1196	1704	2149	1710	159	9.3
					SV	240	517	1407	2276	1400	326	23.3
Positive area[v]	(g·sec)	<0.0001	<0.0001	<0.0001	MC	240	1076	2013	2854	2041	281	13.8
					SV	240	590	1528	2406	1506	355	23.6
Paste-making quality												
Water uptake	(%)	<0.0001	0.001	<0.0001	MC	24	98.0	105.6	113.6	105.8	5.0	4.7
					SV	23	94.9	108.6	122.6	109.0	8.9	8.2
WIRB[u]	(ratio)	<0.0001	0.001	0.001	MC	24	2.94	3.05	3.20	3.06	0.08	2.6
					SV	23	2.65	2.94	3.36	2.98	0.19	6.2
Paste yield	(ratio)	<0.0001	<0.0001	<0.0001	MC	24	45.6	53.7	59.7	54.2	3.4	6.2
					SV	23	28.1	46.7	51.1	44.8	5.8	13.0
Unsweetened paste quality												
L*		<0.0001	0.002	0.073	MC	24	66.9	73.9	75.9	73.2	2.2	3.1
					SV	23	63.7	73.4	75.9	72.5	2.9	4.0
a*		<0.0001	0.072	0.059	MC	24	2.4	3.5	4.7	3.5	0.6	16.4
					SV	23	2.3	3.5	5.2	3.6	0.8	21.5
b*		0.001	<0.0001	0.017	MC	24	9.9	14.3	17.3	14.2	1.7	12.2
					SV	23	13.5	16.2	18.0	16.1	1.4	8.5
C*		0.003	<0.0001	0.011	MC	24	10.9	14.7	17.7	14.6	1.6	10.8
					SV	23	14.2	16.5	18.4	16.5	1.3	8.2
L*/C*		0.011	<0.0001	0.011	MC	24	4.2	5.1	6.1	5.1	0.5	9.8
					SV	23	3.6	4.4	5.2	4.4	0.5	10.5
Whiteness	(score)[t]	<0.0001	0.015	<0.0001	MC	96	1.0	4.0	5.0	4.1	1.2	29.7
					SV	92	1.0	4.0	5.0	3.8	1.1	29.0
Beaniness	(score)	0.003	0.139	0.186	MC	96	1.0	2.0	5.0	2.3	1.3	55.8
					SV	92	1.0	2.0	5.0	1.9	1.0	53.0
Vegetativeness	(score)	0.904	0.645	0.303	MC	96	1.0	2.0	5.0	1.9	1.0	51.8

Table 3.2 (cont'd)

					SV	92	1.0	2.0	5.0	2.0	1.0	51.3
Sweetness	(score)	0.134	0.725	0.625	MC	96	1.0	1.0	3.0	1.2	0.4	35.6
					SV	92	1.0	1.0	5.0	1.3	0.7	57.4
Total flavor intensity	(score)	0.006	0.878	0.048	MC	96	1.0	2.0	5.0	2.3	1.1	48.3
					SV	92	1.0	2.0	5.0	2.3	1.0	45.3
Sweetened paste quality												
L*		<0.0001	0.128	0.748	MC	24	28.8	54.1	69.3	53.1	11.2	21.1
					SV	23	30.2	60.3	69.3	56.9	10.3	18.2
a*		<0.0001	0.045	0.014	MC	24	3.6	5.7	7.1	5.6	1.0	17.2
					SV	23	3.7	5.9	7.6	6.0	1.0	16.1
b*		<0.0001	<0.0001	0.526	MC	24	9.3	23.0	28.5	21.9	4.9	22.4
					SV	23	11.6	25.5	29.3	24.6	4.5	18.1
C*		<0.0001	<0.0001	0.358	MC	24	11.1	23.7	29.3	22.6	4.6	20.5
					SV	23	13.2	26.1	30.2	25.4	4.2	16.5
L*/C*		0.010	0.152	0.720	MC	24	1.8	2.3	3.5	2.4	0.4	16.6
					SV	23	1.8	2.3	2.7	2.2	0.3	11.6
Whiteness	(score)	<0.0001	0.988	0.011	MC	96	1.0	4.0	5.0	3.5	1.4	41.3
					SV	92	1.0	4.0	5.0	3.6	1.2	32.4
Stickiness	(g·sec)	<0.0001	0.065	0.079	MC	24	2979	4204	6910	4378	1123	25.7
					SV	23	2253	4000	5438	4015	843	21.0

[z]S.D.: standard deviation; [y]C.V.: coefficient of variation; [x]MC: Montcalm, [w]SV: Saginaw Valley; [v]Peak Positive Force and Positive Area: Soaked hardness measurements; [u]WIRB: Weight Increase Rate by Boiling. [t]Score: 1-5 scale with 5 being the strongest.

Table 3.3 Mean values for agronomic and paste quality traits of six bean genotypes grown in the two environments in Michigan.

	(unit)	Environment	Otebo Hime	Otebo Samurai	Navy Alpena	Great northern Powderhorn	White kidney Snowdon	Cranberry CR1502-4
Seed quality								
100 seed weight	(g)	MC[z]	27.0 c	23.5 d	16.9 e	33.3 b	63.9 a	63.6 a
		SV[y]	25.5 d	24.1 d	17.2 e	32.9 c	52.2 a	46.0 b
Seed coat	(%)	MC	7.6 a	7.5 a	6.9 b	7.9 a	6.9 b	6.9 b
		SV	8.6 a	7.6 cb	7.0 c	7.9 ab	8.0 ab	8.6 a
Peak Positive Force[x]	(g)	MC	1875 a	1724 b	1747 b	1703 b	1599 c	1611 c
		SV	1272 dc	1454 b	1382 bc	1752 a	1105 d	1437 bc
Positive area[x]	(g·sec)	MC	2320 a	1979 bc	1779 d	2214 a	1892 dc	2059 b
		SV	1421 dc	1505 bc	1299 d	1945 a	1255 d	1612 b
Paste quality								
Water uptake	(%)	MC	107.9 abc	101.7 dc	103.6 dbc	111.4 a	109.2 ab	100.8 d
		SV	116.2 ab	97.2 c	100.3 c	110.5 b	119.5 a	111.4 ab
WIRB[w]	(ratio)	MC	3.17 a	3.07 ab	2.98 b	3.05 ab	3.05 ab	3.08 ab
		SV	3.06 ab	3.26 a	2.89 cb	2.79 c	2.92 cb	2.91 cb
Paste yield	(ratio)	MC	52.3 cb	53.9 cb	49.9 c	54.8 b	59.1 a	55.1 ab
		SV	45.3 a	49.0 a	44.3 a	45.9 a	48.0 a	33.2 b
Unsweetened paste quality								
L*		MC	73.9 a	73.0 a	74.2 a	74.4 a	74.8 a	68.9 b
		SV	73.4 a	72.3 a	72.7 a	74.6 a	74.3 a	65.7 b
a*		MC	3.3 cb	3.8 ab	3.7 ab	3.2 cb	2.8 c	4.2 a
		SV	3.5 bc	4.1 b	3.8 bc	2.7 d	3.0 dc	5.0 a
b*		MC	14.1 b	13.0 b	13.5 b	16.8 a	15.0 ab	12.7 b
		SV	16.4 NS[v]	14.9	15.3	16.4	16.8	17.1
C*		MC	14.5 b	13.6 b	14.0 b	17.1 a	15.3 ab	13.4 b
		SV	16.8 NS	15.4	15.7	16.6	17.1	17.8
L*/C*		MC	5.1 ab	5.4 a	5.3 a	4.4 b	4.9 ab	5.2 ab
		SV	4.4 ab	4.7 a	4.6 ab	4.5 ab	4.4 ab	3.7 b
Whiteness	(score)[u]	MC	4.9 a	4.5 ab	4.5 ab	4.2 b	4.7 ab	1.7 c
		SV	3.8 b	3.8 b	4.1 ab	4.6 a	4.2 ab	1.6 c
Beaniness	(score)	MC	1.5 c	1.9 abc	1.9 bc	2.3 abc	2.8 ab	3.1 a
		SV	2.0 NS	1.6	1.8	1.9	2.3	2.2

Table 3.3 (cont'd)

Vegetativeness	(score)	MC	1.7 NS	2.1	1.6	2.1	2.1	2.1
		SV	2.0 NS	1.7	2.4	1.9	2.1	2.0
Sweetness	(score)	MC	1.2 NS	1.3	1.2	1.2	1.2	1.1
		SV	1.2 NS	1.6	1.2	1.4	1.1	1.0
Total flavor intensity	(score)	MC	1.7 b	2.1 b	1.9 b	2.3 b	2.4 ab	3.4 a
		SV	2.2 NS	2.0	2.3	2.4	2.5	2.3
Sweetened paste quality								
L*		MC	52.7 a	52.0 a	60.2 a	54.9 a	64.9 a	33.7 b
		SV	60.0 a	56.0 a	59.9 a	60.2 a	63.9 a	35.6 b
a*		MC	5.3 ab	5.7 a	5.8 a	6.5 a	4.0 b	6.1 a
		SV	6.8 a	6.3 ab	6.0 ab	5.3 ab	4.9 b	6.6 ab
b*		MC	22.9 b	21.1 b	22.5 b	27.8 a	24.1 ab	12.9 c
		SV	27.1 ab	23.3 b	24.3 b	28.6 a	26.7 ab	15.0 c
C*		MC	23.5 b	21.9 b	23.2 b	28.5 a	24.4 b	14.3 c
		SV	28.0 ab	24.2 b	25.1 ab	29.1 a	27.1 ab	16.4 c
L*/C*		MC	2.2 ab	2.4 ab	2.6 ab	1.9 b	2.7 a	2.4 ab
		SV	2.2 NS	2.3	2.4	2.1	2.4	2.2
Whiteness	(score)	MC	4.3 ab	3.5 b	3.7 ab	3.8 ab	4.5 a	1.2 c
		SV	3.3 b	3.4 b	3.9 ab	4.4 a	4.4 a	1.7 c
Stickiness	(g·sec)	MC	4116 ab	3836 b	3738 b	6094 a	4459 ab	4023 b
		SV	4790 a	3030 c	3947 abc	4792 a	4245 ab	3065 bc

Means with the same letter in the same row are not different (α=0.05). [z]MC: Montcalm; [y]SV: Saginaw Valley; [x]Peak Positive Force and Positive Area: hardness measurements of soaked seeds; [w]WIRB: Weight Increase Rate by Boiling; [v]NS: Not significant. [u]Score: 1-5 scale with 5 being the strongest.

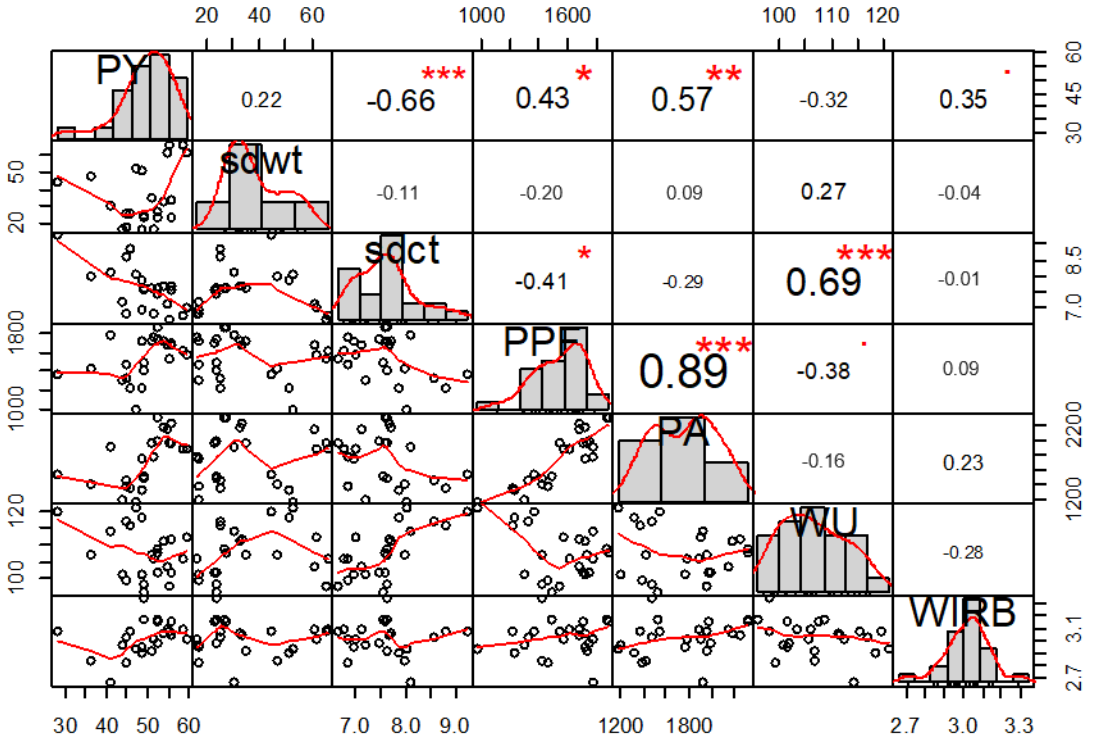
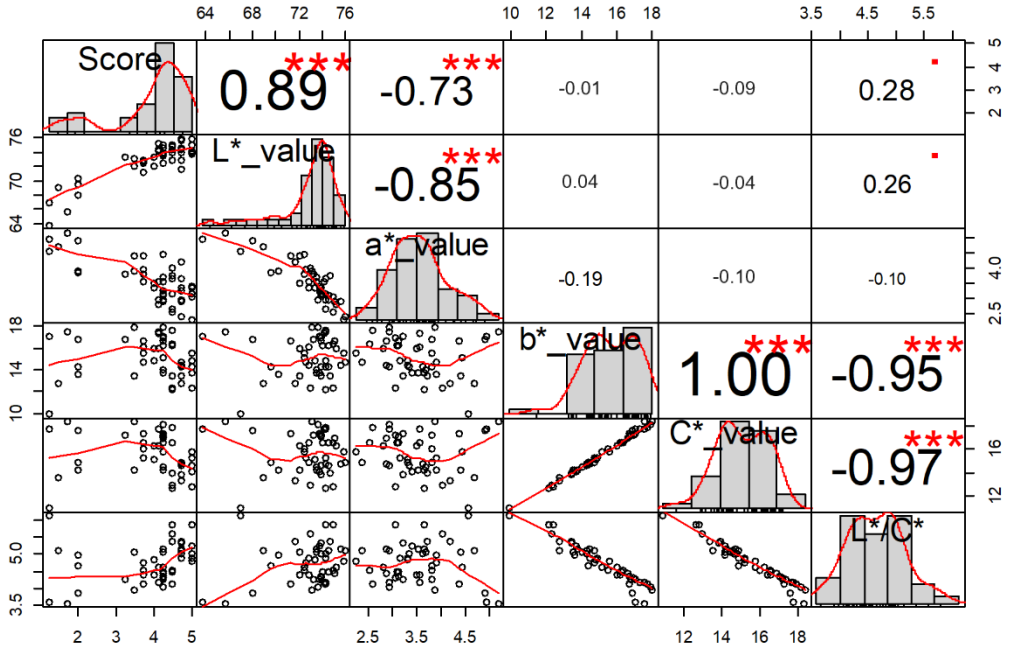


Figure 3.1 The correlation coefficients and scatter plots of paste-making and cooking quality traits with datapoints from the two environments combined. PY: paste yield; sdwt: 100 seed-weight; sdct: seed coat percentage; PPF: peak positive force of soaked seeds; PA: positive area of soaked seeds; WU: water uptake; WIRB: weight increase rate by boiling. ***: *p*-value<0.001; **: *p*-value<0.01; *: *p*-value<0.05.

A



B

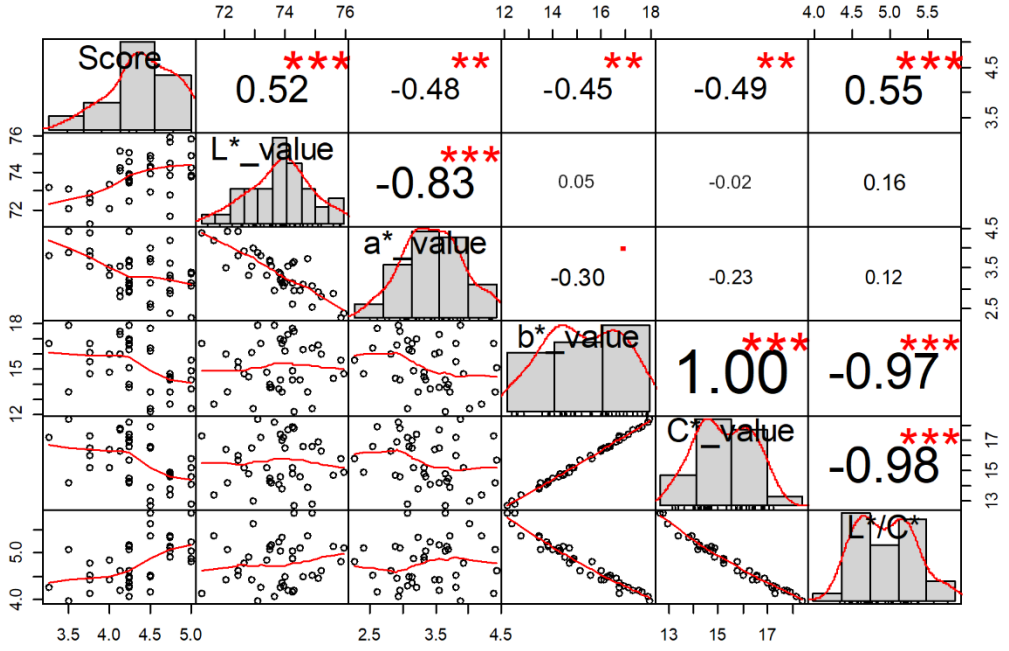
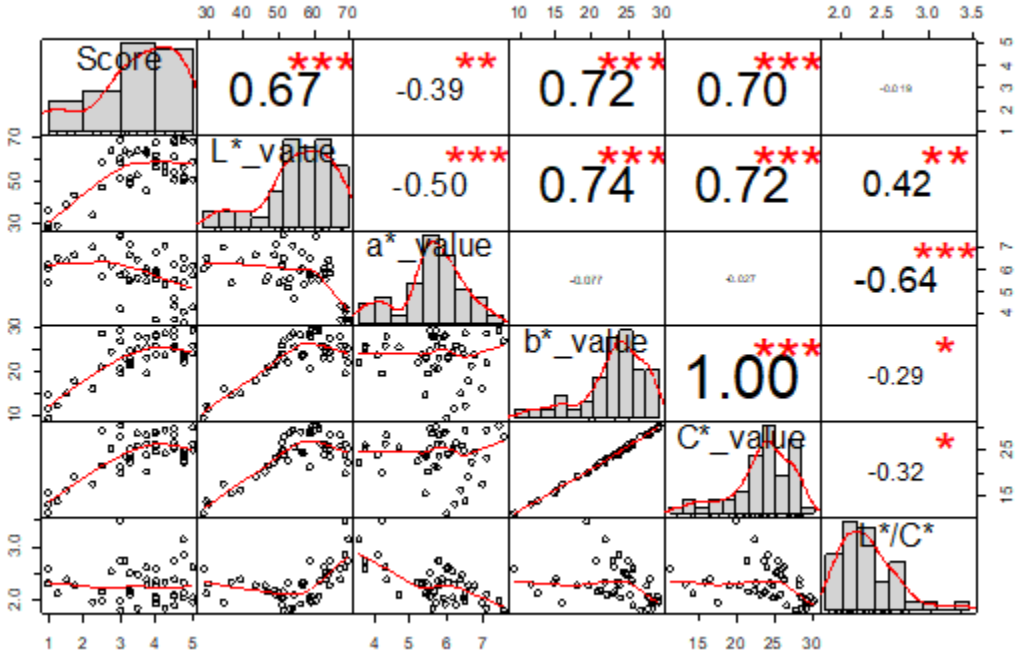


Figure 3.2 The correlation coefficients and scatter plots of whiteness scores evaluated by assessors and color values of unsweetened paste. Score: whiteness score rated by assessors. L*, a*, b*: color values obtained by image analysis; C*: chroma; L*/C*: the ratio of L* and C*. **A**: all data points (n=47) used; **B**: data points except CR1502-4 (n=40). ***: *p*-value<0.001; **: *p*-value<0.01; *: *p*-value<0.05.

A



B

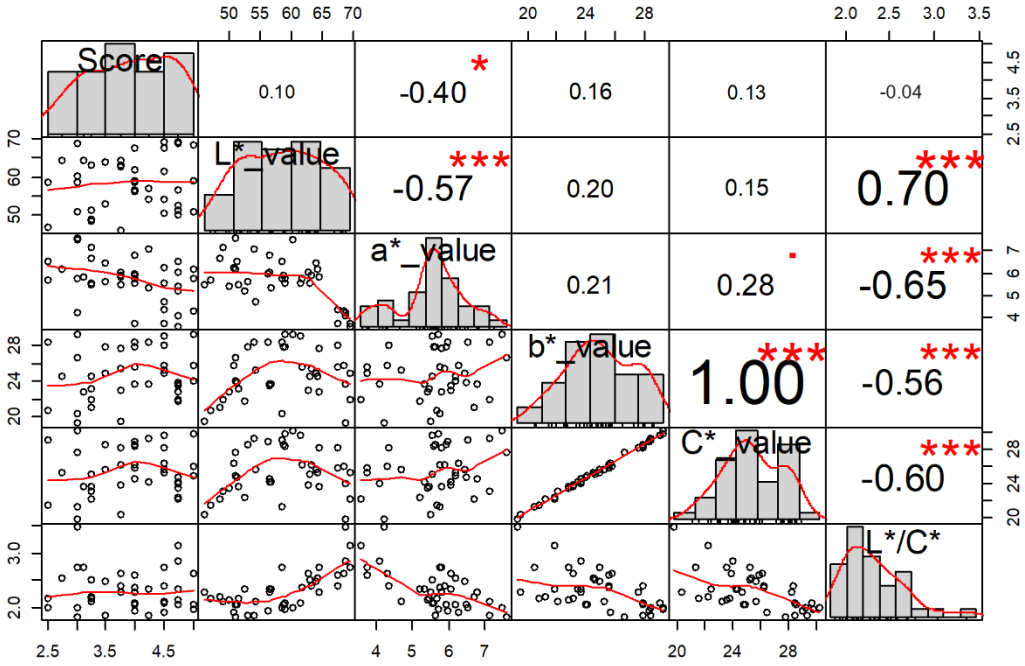


Figure 3.3 The correlation coefficients and scatter plots of whiteness scores evaluated by assessors and color values of sweetened paste. Score: whiteness score rated by assessors. L*, a*, b*: color values obtained by image analysis; C*: chroma; L*/C*: the ratio of L* and C*. **A**: all data points (n=47) used; **B**: data points except CR1502-4 (n=40). ***: *p*-value<0.001; **: *p*-value<0.01; *: *p*-value<0.05.

REFERENCES

REFERENCES

Abramoff, M., Magalhaes, P., & Ram, S. (2004). Image processing with ImageJ. *Biophotonics International*, *11*(7), 36–42.

CIE International Commission on Illumination. (1978). Recommendations on Uniform Color Spaces, Color-Difference Equations, Psychometric Color Terms. In *Supplement No.2 to CIE Publication No. 15, Colorimetry, 1971*.

Desrochers, N., & Brauer, P. (2001). Legume promotion in counselling: an e-mail survey of dietitians. *Canadian Journal of Dietetic Practice and Research*, *62*(4), 193–198.

Dorrell, D. G., & Adams, M. W. (1969). Effect of some seed characteristics on mechanically induced seedcoat damage in navy beans (*Phaseolus vulgaris* L.). *Agronomy Journal*, *61*(5), 672–673. https://doi.org/10.2134/agronj1969.00021962006100050006x

Hayat, I., Ahmad, A., Masud, T., Ahmed, A., & Bashir, S. (2014). Nutritional and health perspectives of beans (*Phaseolus vulgaris* L.): an overview. *Critical Reviews in Food Science and Nutrition*, *54*(February 2013), 580–592. https://doi.org/10.1080/10408398.2011.596639

Kamiya, M., Takeuchi, T., & Kusume, T. (2004). Identification of a white common bean cultivar "Yukitebo" by DNA polymorphisms using PCR amplification [In Japanese]. *Breeding Research*, *6*, 29–32.

Kato, J. (2000). Studies on characteristics for food processing of Adzuki beans and common beans, and factors for their variation [English abstract] [Iwate University]. In *Hokkaido Central Agricultural Experiment Station Report* (Vol. 95). https://www.hro.or.jp/list/agricultural/center/kankoubutsu/houkoku/houkoku5.htm

Kelly, J. D., & Cichy, K. A. (2012). Dry bean breeding and production technologies. In M. Siddiq & M. A. Uebersax (Eds.), *Dry beans and pulses production, processing and nutrition* (pp. 23–54). https://doi.org/10.1002/9781118448298.ch2

Kelly, J. D., Varner, G. V., Cichy, K. A., & Wright, E. M. (2012). Registration of 'Snowdon' white kidney bean. *Journal of Plant Registrations*, *6*(3), 238–242. https://doi.org/10.3198/jpr2012.03.0146crc

Kelly, J. D., Varner, G. V., Cichy, K. A., & Wright, E. M. (2014). Registration of 'Powderhorn' great northern bean. *Journal of Plant Registrations*, *8*(1), 1–4. https://doi.org/10.3198/jpr2013.05.0020crc

Kelly, J. D., Varner, G. V., Cichy, K. A., & Wright, E. M. (2015). Registration of 'Alpena' navy bean. *Journal of Plant Registrations*, *9*(1), 10–14. https://doi.org/10.3198/jpr2014.04.0025crc

Kelly, J. D., Varner, G. V., Hooper, S., Cichy, K. A., & Wright, E. M. (2016). Registration of

‘Samurai’ Otebo Bean. *Journal of Plant Registrations*, *10*, 109–114. https://doi.org/10.3198/jpr2015.09.0051crc

Kelly, J. D., Varner, G. V., Roman, B., & Long, B. (2009). Registration of ‘Fuji’ Otebo bean. *Journal of Plant Registrations*, *3*(3), 223–225. https://doi.org/10.3198/jpr2008.12.0733crc

Komiyama, S. (2013). The mechanism of darkening of Tebo Ann and a method for measuring color changes [In Japanese]. *Hokkaido Central Agricultural Experiment Station Report*, 1–5.

Komiyama, S., & Kato, J. (2004). Evaluation method for texture (stickiness) of white bean paste made of common beans. *Hokkaido Central Agricultural Experiment Station Report*, *86*, 65–72.

Mendoza, F. A., Kelly, J. D., & Cichy, K. A. (2017). Automated prediction of sensory scores for color and appearance in canned black beans (*Phaseolus vulgaris* L.) using machine vision. *International Journal of Food Properties*, *20*(1), 83–99. https://doi.org/10.1080/10942912.2015.1136939

Messina, V. (2014). Nutritional and health benefits of dried beans. *The American Journal of Clinical Nutrition*, *100*(1), 437S-442S. https://doi.org/10.3945/ajcn.113.071472.2

Miura, T., Narikawa, T., Ushirogi, T., & Inuzuka, T. (1977). New common bean variety “Hime-tebo” [English abstract]. *1Hokkaido Central Agricultural Experiment Station Report*, *38*, 83–91.

Mkanda, A. V, Minnaar, A., & de Kock, H. L. (2007). Relating consumer preferences to sensory and physicochemical properties of dry beans (*Phaseolus vulgaris*). *Journal of the Science of Food and Agriculture*, *87*(15), 2868–2879. https://doi.org/10.1002/jsfa.3046

Ogle, D. H., Wheeler, P., & Dinno, A. (2019). *FSA: Fisheries Stock Analysis* (R package version 0.8.24). https://github.com/droglenc/FSA

Okuyama, M., Ebe, S., Sato, H., Mikami, K., Murata, K., Shimada, H., Ogawa, T., & Shimosaka, M. (2008). A new common bean variety “Kinu-tebo” [English abstract]. *Hokkaido Central Agricultural Experiment Station Report*, *92*, 13–27.

Ovando-Martínez, M., Bello-Pérez, L. A., Whitney, K., Osorio-Díaz, P., & Simsek, S. (2011). Starch characteristics of bean (*Phaseolus vulgaris* L.) grown in different localities. *Carbohydrate Polymers*, *85*(1), 54–64. https://doi.org/10.1016/j.carbpol.2011.01.043

Peterson, B. G., & Carl, P. (2019). *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis* (R package version 1.5.3). https://cran.r-project.org/package=PerformanceAnalytics

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Rivera, A., Fenero, D., Almirall, A., Ferreira, J. J., Simó, J., Plans, M., Romero del Castillo, R., &

Casañas, F. (2013). Variability in sensory attributes in common bean (*Phaseolus vulgaris* L.): a first survey in the Iberian secondary diversity center. *Genetic Resources and Crop Evolution*, *60*(6), 1885–1898. https://doi.org/10.1007/s10722-013-9963-6

Sadohara, R. (2019). Quality characteristics of bean paste as a confectionery ingredient and recent breeding efforts of common beans in Japan. *Journal of the Science of Food and Agriculture*, *100*, 10–15. https://doi.org/10.1002/jsfa.10013

Shinada, Y., Iida, S., Chiba, I., Hara, M., Sato, H., & Nakano, M. (1994). A new common bean variety "Yuki-tebo" [English abstract]. *Hokkaido Central Agricultural Experiment Station Report*, *66*, 25–34.

Shiota, Y., & Miyata, Y. (1976). Studies on "Ann" (bean jam) (part 4) [English abstract]. *Journal of Home Economics of Japan*, *27*(3), 180–185.

Smith, D. K., Anne Riddle, L., Kerr, S., Atterberry, K., Lanigan, J., Miles, C., Riddle, L. A., Kerr, S., Atterberry, K., Lanigan, J., & Miles, C. (2016). Barriers and Opportunities to Serving Pulses in School Meals in Washington Schools. *The Journal of Child Nutrition and Management*, *12*(1). https://schoolnutrition.org/uploadedFiles/5_News_and_Publications/4_The_Journal_of_Child_Nutrition_and_Management/Spring_2016/11-BarriersandOpportunitiestoServingPulses.pdf

Tazawa, A., Sato, H., Shinada, H., Aoyama, S., Fujita, S., Murata, K., Matsukawa, I., & Hasegawa, N. (2015). A new Dinagon-brand Adzuki bean variety "Homare-dainagon" with soil-borne disease resistance and high processing adaptability. *Hokkaido Central Agricultural Experiment Station Report*, *99*, 1–11.

Tyler, R., Wang, N., & Han, J. (2017). Composition, nutritional value, functionality, processing, and novel food uses of pulses and pulse ingredients. *Cereal Chemistry Journal*, *94*(1), 1. https://doi.org/10.1094/CCHEM-12-16-0500-R

US Department of Agriculture Economic Research Service. (2017). *U.S. per capita use of fresh and processing vegetables, dry pulse crops, and potatoes; cash receipts; U.S. vegetable trade*. Vegetables and Pulses Yearbook Tables. https://www.ers.usda.gov/data-products/vegetables-and-pulses-data/vegetables-and-pulses-yearbook-tables/

Wu, D., & Sun, D.-W. (2013). Colour measurements by computer vision for food quality control – A review. *Trends in Food Science & Technology*, *29*(1), 5–20. https://doi.org/10.1016/j.tifs.2012.08.004

Yam, K. L., & Papadakis, S. E. (2004). A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering*, *61*(1), 137–142. https://doi.org/10.1016/S0260-8774(03)00195-X

CHAPTER 4:

FOOD INDUSTRY VIEWS ON PULSE FLOURS – PERCEIVED INTRINSIC AND EXTRINSIC CHALLENGES FOR PULSE FLOUR UTILIZATION

Food Industry Views on Pulse Flours – Perceived Intrinsic And Extrinsic Challenges For Pulse Flour Utilization

Rie Sadohara [1], Donna M. Winham [2], Karen A. Cichy [1,3]

1. The Department of Plant, Soil and Microbial Sciences, Michigan State University, 1066 Bogue St. East Lansing, MI. 48824, USA
2. The Department of Food Science and Human Nutrition, Iowa State Unviersity, 38 MacKay 2302 Osborn Dr. Ames, IA 50011, USA
3. USDA-ARS, Sugarbeet and Bean Research Unit, 1066 Bogue St. East Lansing, MI. 48824, USA

Abstract

Pulses such as beans, chickpeas, peas, and lentils offer numerous health and nutritional benefits, and milled pulse flours will increase their versatility and may increase pulse consumption. Beans are the most produced pulse crop in the U.S. and have a potential to serve as a milled ingredient for large-scale food manufacturing; however, their use as flour is limited. Pulse seed traits that the food industry wants to be improved for flour purposes are largely unknown. This study aimed to understand the food industry's needs and barriers for pulse flour utilization via an electronic survey targeting employees at food companies that manufacture products typically made from wheat. Survey questions asked about intrinsic factors of pulse flours such as end product quality characteristics they find satisfactory or challenging, as well as extrinsic factors including market demand, pulse flour supply, and cost. Twenty-one current, previous, or considered users of pulse flours and 54 non-users completed the survey. Ten of the users indicated that there were challenges utilizing pulse flour while five indicated there were not. The three most selected challenges of end product qualities were flavor, texture, and dough handling properties. The optimum attributes

would need to be identified for specific combinations of pulse flour and product type with a larger sample size. Over half of respondents were not familiar with bean flour, especially non-users, which warrants increased awareness of bean flour. Market demand for improved nutritional products may be the most important extrinsic factor rather than supply or cost.

Introduction

Pulses are starchy seeds of leguminous crops and can serve as a source of protein, dietary fiber, vitamins, and minerals, and provide protective effects against cardiometabolic diseases and certain types of cancer (Hall et al., 2017; Mudryj et al., 2014). The major pulse types produced in the U.S. include dry beans, dry peas, chickpeas, and lentils (FAOSTAT, 2021). Due to their high nutritive value, pulse consumption is encouraged by the Dietary Guidelines for Americans (US Department of Agriculture & US Department of Health and Human Services, 2015). However, pulse consumption is low among consumers; only 7.9% of U.S. adults consumed pulses including dry beans on a given day (Mitchell et al., 2009). Barriers for pulse utilization include limited time for meal preparation and lack of knowledge on pulse cooking (Palmer et al., 2018; Winham et al., 2019). Milled pulses (pulse flours) offer a broader range of applications such as baked goods, snacks, and pasta, which are ready to eat or quick to prepare. The high levels of protein and dietary fiber (15-30% each) in pulses make pulse flours an ideal ingredient to nutritionally improve wheat-flour based products (Abu-Ghannam & Gowen, 2011; Hall et al., 2017). Alternatives to wheat flour are gaining attention, especially rice and chickpea flours, due to consumers' increasing interest in the nutritional and health benefits of pulse flours (American Pulse Association, 2020; Teodoro, 2017). In fact, 3,449 new pulse-based products were launched in the U.S. from 2006 to

2016, and more than half of the products were made from peas or pea protein isolate (Zarrouki, 2017).

Numerous studies have been conducted to investigate the suitability of pulse flours for various food applications (Malcolmson & Han, 2019; Maskus, 2010; Sozer et al., 2017). Pulse flour utilization brings some challenges such as off- and beany flavors, texture, and functional properties different from wheat flour due to the lack of gluten (Marti & Pagani, 2013; Simons & Hall, 2018; Sozer et al., 2017). Thakur et al. (2019) compared pulse and wheat flours with established quality parameters for milling. They argued that specification of pulse flours and seed quality parameters for pulse milling need to be defined for consistency; and that the effect of genotype, environment, storage, processing, and milling on those quality parameters, as well as flour components and their interactions need to be studied (Thakur et al., 2019). Performance can differ depending on pulse flour and application types, and flours can be characterized by many attributes such as dehulling efficiency, physicochemical properties, thermal behavior, and nutrient composition (Farooq & Boye, 2011; Martínez, 2015; Maskus, 2010; Wood & Malcolmson, 2011). End-product qualities such as appearance, aroma, and flavor are also important to consider for consumer acceptability.

The perspectives of end-users on pulse flours are scarce in the literature. One case study was conducted (Lascialfari et al., 2019) with European and multinational food manufacturers of pulse-based products, which identified important factors for innovation in pulse-based product development. These were grouped into three categories: technical aspects, market trends, and ingredient availability. Technical aspects include the functionality and behavior of pulse flours when processed, and difference between regular wheat and pulse flours are noted as expected. Market trends are strong drivers for their product development, and depend on time and country;

consumers' conservative attitude in Europe was deemed to be a challenge, while pulse-based pasta was more accepted in the U.S. Finally, ingredient availability was a concern, especially to firms with large production capacity. Since this study was conducted almost exclusively with firms in Europe, a similar study with U.S. food manufacturers would facilitate understanding of the challenges with pulse flours more specific to the U.S. Important traits depend on application types, so optimum pulse flour traits need to be specified for each type of application.

Common bean (*Phaseolus vulgaris*) is the most produced pulse crop globally and in the U.S. Approximately one million tons of dry beans are produced annually and consumed primarily in the form of canned beans in the U.S. (Parr & Lucier, 2020) However, beans can also be milled into flour and can be used for food products that are typically made from wheat flour. Bean flour has been tested for various food applications on a laboratory scale such as quick bread (Alani et al., 1989; Dryer et al., 1982), tortillas (Mora-Avilés et al., 2007), pasta (Gallegos-Infante et al., 2010; Hooper et al., 2019), cookies (Simons & Hall, 2018), biscuits (Sparvoli et al., 2016), cakes, crackers, and meat analogues (Szczygiel et al., 2017), with promising results. Bean flour has been commercially produced and used, but its market presence remains rather small. Aroma, appearance, texture, and beany flavor, likely caused by lipid oxidation, may be a challenge (Szczygiel et al., 2017), not only for bean flour but also for other types of pulse flours (Roland et al., 2017; Thakur et al., 2019). Among the aforementioned 3,449 pulse-based products launched, bean flour was not in the top ten ingredients (Zarrouki, 2017). This suggests that there is a need to improve beans for use as flour, but which traits do not meet users' expectation are not clear.

Due to the large number of products that utilize pulse flours, interests and needs for pulse flours by the food industry is an important missing piece of information for developing pulse

varieties for flour purposes. This survey study aimed to identify needs for and the barriers limiting adoption of pulse flours by the U.S. food industry.

Materials and Methods

Ethics approval

This survey study was deemed exempt under 45 CFR 46. 104(d) 2ii by the Institutional Review Board at Michigan State University (Study ID: STUDY00002778). Respondents gave consent to participate in the study by clicking ”Yes, I agree to participate in this survey”. No compensation for participation was offered or provided.

Survey development

Survey questions were developed through interviews with academic researchers and bean industry personnel involved in bean or pulse flour research and promotional activities of dry beans. The interview questions included sections about corporate research and development pipeline and their experience with and impression of pulse flours. Sixteen convenience samples of flour or milling industry professionals were invited to participate in an interview via teleconference. Two answered the interview questions by email, and three provided their general opinions about pulse flours instead of answering the interview questions. Survey questions were developed based on these responses and were improved by information presented at the Pulse Science and Technology Forum (Toronto, November 5-7, 2019), the Pulse Flour Summit (Minneapolis, MN, March 10-12, 2020), and iterative discussions and edits by all of the authors. Further, the survey was reviewed by 12 academic researchers and industry personnel who did not participate in the interview and was refined based on their feedback.

Survey questions for all participants

User status, pulse flour type, and product types

The survey asked participants their pulse flour use status (currently using, previously used, have considered or currently testing, or never used). Current, previous, and considered (CPC) users were asked for their pulse flour type and product type that they are/were/have considered using. Non-users were asked the flour and product types that they would be interested in using. If they selected multiple pulse flour types, they were asked to select the one most important to them. "Do not know which pulse flour type is best for my production purpose" was also provided as an option to non-users. The pulse flour types included bean, chickpea, pea, lentil, faba, and other; product type included yeast breads, cakes, quick breads, pastries, cookies, spaghetti, Asian noodles, snack foods, thickening agents, coatings and breadings, functional protein products, and plant-based meat alternatives. The pulse flour type and product type questions required answers to proceed.

Job titles and business types

Questions about the participants and their companies were asked: their primary job titles and business type (single location, headquarters, or branch).

Information, collaboration, and impression on bean flour

Respondents were asked if they have ever looked for technical information on the pulse flour they selected, and if they have, follow-up questions were asked about the adequacy of information they found and the sources of information they used. They were asked whether they would be interested providing their opinions to university researchers and plant breeders about how pulse flours can

be improved. A multiple-choice question was asked their impression about bean flour, a pulse type that is most produced in the U.S. but less utilized as flour. Multiple answers were allowed.

Questions for CPC users

Satisfactory and challenging characteristics of pulse-flour based products

CPC users were asked about product characteristics made from pulse flours that they found satisfactory or challenging. Multiple answers were allowed. The reason for currently using or quitting using the pulse flour they selected was asked with multiple choices.

Variability, specifications, and supply

Participants' opinions were asked with choices "Yes", "Sometimes", "No", and "Do not know" regarding variability in pulse flour quality by batch or supplier, their importance in their product quality, universal specification of pulse flours, gluten contamination, and lectins. Questions were asked with choices "Yes", "No", "Do not know" about cost and supply of the pulse flour they selected. The change in availability of the pulse flour due to the COVID-19 pandemic since March 2020 were also asked. All questions were posed with an optional comment box for text entry.

Target population and survey distribution

Qualtrics (Qualtrics, Provo, UT, USA) was used to develop and distribute a survey regarding pulse flour use. A group of Standard Industrial Classification (SIC) codes were identified for food products typically made from wheat flour: 2045, 2051, 2052, 2053, 2096, and 2098 (**Table 4.1**), and 10,292 email addresses were obtained from InfoUSA Marketing, Inc. (St. Louis, MO, USA) of food manufacturing firm employees with one or more of the selected SIC codes. A unique

survey link was generated for each individual to enable response tracking, and they were invited to participate via email. An additional 16 individuals who were suggested by the initial invitees or other sources were invited via email with a unique survey link. A second reminder invitation was sent to 6,741 individuals who had not started the survey or who had started but not finished and had a job title considered to be relevant to this study (**Table S4.12**). A reminder email was sent 2-7 days after the first and second distributions.

An anonymous URL (weblink) for the survey was generated to allow anyone with the link access to the survey. The weblink was linked to the survey with the same questions regarding pulse flours, but had additional voluntary questions asking participants' first names, business name and zip code, and the type of business (single location, headquarters, branch, or other). The weblink was distributed via U.S. Dry Bean Council, Michigan Bean Commission, USA Dry Pea and Lentil Council, faculty members at the Department of Food Science and Human Nutrition of Iowa State University including the Departmental External Advisory Committee, Northharvest, Institute of Food Technologists (IFT) Iowa, American Pulse Association, faculty members at the Department of Food Science and Human Nutrition of Michigan State University, and IFT Great Lakes Section. The survey was also distributed with the weblink on authors' accounts on Twitter and LinkedIn.

Results and discussion

Eighty-four responses were collected in total. Nine responses from outside the U.S. or irrelevant businesses were removed, resulting in 75 valid responses, of which six were collected via the weblink. Respondents were categorized into four pulse flour user types: eight current, four previous, nine considered, and 54 non-users. The most common role of the respondents was management or owner (29), followed by research and development (R&D, 18) (**Table2**). There

were 30 respondents in R&D or production/chef, who are most likely to be knowledgeable about the behavior of pulse flours when utilized in food products. A half of current and 67% of considered users were in R&D, whereas 25% and 13% of previous and non-users were in R&D, respectively.

Pulse flour and product types

Participants were asked the pulse flour type they are currently, previously, considered, or would be interested in using, and the product type they would like to make with the pulse flour they selected (**Table 4.3**). Chickpea and pea flours were by far the most selected by all types of users rather than bean, faba, and lentil flours, whereas 70% of non-users did not know which pulse type was best for their needs. Thirteen percent of non-users explicitly stated that they were not interested in using any pulse flours. Over half of considered users selected chickpea, whereas 63% of current and 50% of previous users selected pea flour. Among all the respondents, yeast breads were the most selected products they are currently producing, previously produced, have considered, or would be interested in using pulse flour for, followed by cookies and quick breads (**Table 4.4**). Twenty percent of non-users explicitly stated that they were not interested in using pulse flours for any kinds of products. Though chickpea and pea were the most selected pulse type (**Table 4.3**), the type of products using the pulse flours selected varied widely (**Table 4.5**). However, yeast breads and cookies (9 times each) were relatively popular among those who did not know which type of pulse flour is best for their needs.

Information

Thirty-one percent (23 out of 75) of all respondents indicated that they have looked for pulse flour technical information (**Table 4.6**). More CPC users had looked for information on pulse flours than non-users, and 89% of non-users never looked for technical information on pulse flours. Among those who had looked for information, 61% (14 out of 23) said they were able to find sufficient information.

Participants who had looked for information on pulse flours were asked to select all sources of information they used. For-profit researchers and industry associations such as the Institute of Food Technologists were the most selected source of information (**Figure 4.1**). Non-profit organizations such as the American Pulse Association and academic textbooks and researchers were the second most selected. Open-access journals were more popular than subscribed and pay-per-view journals. Those who selected "Other" commented that information was obtained from their suppliers (3) and from the Internet (1). From these results, academic and for-profit organizations and communication between researchers seemed to be effective channels that are likely to reach pulse flour end-users.

Collaboration with pulse breeders

Participants were asked if they would be interested in providing opinions to university researchers and plant breeders about how pulse flours can be improved. Nine selected "Yes", six "Other", and 59 "No". Those who selected "Other" left comments that they do not know enough about pulse flours to provide useful feedback or that someone in the company might be willing to. Business type information was available for 72 of the 74 who answered this question. Seven out of the 54 participants at a single location were willing to collaborate, while two out of the 18 at large firms

that have headquarters, subsidiary and/or branches were willing to (**Figure 4.2**). Though a small number, nine individuals were willing to collaborate with pulse breeders, which will be useful in developing varieties that meet the end-users' needs and will increase pulse flour adoption. For example, an industry partner was involved in the lab-scale evaluation of pulse flours for crackers, and a selected pulse type (chickpea) was further tested on a commercial scale at the company (Han et al., 2010).

Impression on bean flour

To gain insights on how beans could be improved for flour purposes, participants were asked for their impressions on bean flour. Multiple answers were allowed. "Do not know about bean flour" was the most selected impression on bean flour, especially by non-users (57%). However, 25-33% of CPC users also did not know about bean flour (**Table 4.7**). These results justify increasing promotional activities for bean flour. Over 60% of current and considered users selected "Flavor is a challenge". All user types selected "Market demand for bean flour is low". Flavor and functionality were selected as a challenge 22 and 21 times respectively, whereas gluten contamination and lectin concerns were only selected 6 times in total, indicating that participants are more concerned about culinary and processing qualities than safety of bean flour. A comment mentioned the expected low production volume of bean flour: "Volume limitations limit its commercial usage". To our knowledge, the proportion of beans used as flour is unknown.

Pulse flour and product types of CPC users

There were six users who are currently producing, were previously producing, or have considered producing yeast breads, five snack foods, four meat alternatives, three quick breads, followed by

cakes, cookies, and thickening agents prepared by one each (**Table 4.8**). For the eight current users, the most selected reason for using pulse flours was driven by marketing and trends, followed by protein content and functional characteristics of the pulse flour (**Table 4.9**). The reason for discontinuing pulse flour by the four previous users were all "Low market demand for the product made with the pulse flour", and no other choices regarding flour performance, supply, or allergen contamination were selected as the reason for discontinuation. These results indicate that market demand is an important determinant on whether or not to use pulse flour. Likewise, low market demand ranked fourth for bean flour impression (**Table 4.7**).

Satisfactory and challenging characteristics of product quality

The CPC users were asked to select satisfactory traits of pulse flour. Multiple answers were allowed. Texture, appearance, and uniformity were the top three satisfactory characteristics (**Figure 4.3A**). Five current users were satisfied with product color with pulse flours, while only two previous or considered users were. This may indicate the importance of product color for commercial use of pulse flours. Flavor was selected as satisfactory by all user types, which was surprising considering the deemed challenges regarding the 'beany' flavor of pulse flours (Bresciani & Marti, 2019; Roland et al., 2017). This suggests that some food manufacturers know ways to adjust and optimize flavor. Beany flavor could or could not be problematic depending on seasoning used and on the incorporation rate of pulse flours if composite flour is used (Dabija et al., 2017; Szczygiel et al., 2017). Unpleasant flavor is also tackled by a plant breeding approach; a commercially available pea protein ingredient is "produced from U.S. yellow pea varieties, specifically selected to minimize the off-flavors normally associated to pulses" (quote from Cargill, 2019).

Eighteen out of the 21 users answered the question “Were there challenges when using the pulse flour?”. Ten users indicated that they have encountered challenge(s) when using their pulse flour, whereas five indicated they did not (**Table 4.10**). The combination of pulse flour and product type is shown in **Table 4.11**. Among the 10 respondents who reported that they had challenges producing yeast breads (5), snack foods (2), meat alternatives (2), and cookies (1) with chickpea, pea, faba, and bean flours. In contrast, five respondents said there were no challenges for snack foods (2), quick breads (2), and thickening agents (1) with chickpea, pea, and bean flours. It was noteworthy that four respondents used bean and chickpea flours for snack foods, and two reported that there were challenges, whereas the other two reported there were not. Though the number of respondents was small, it was intriguing that some of the users in all the user statuses indicated that there were no challenges and that their product was satisfactory, indicating that it is possible to produce satisfactory quality products for some combinations of pulse flour and product type.

Yeast breads were the main product type that pulse flours were used for and came with challenges most often. This high count may be partly because yeast breads were the most selected products participants have used or considered using pulse flours for (**Table 4.4**). However, the challenges that yeast bread producers reported are useful information because gluten formation plays a vital role in yeast bread quality, and pulse flours, which are gluten-free, often presents challenges in terms of loaf volume and texture when made into breads (Melini et al., 2017). This survey was designed to characterize the combination of pulse flour and end product as well as the production conditions so that conclusions could be drawn for specific pulse flour type and application; nevertheless, this was not possible due to a small sample size. Thus, an alternative approach for pulse breeders might be to collaborate with an industry partner to provide feedback on pulse varieties for flours under a confidentiality agreement.

The 10 respondents who reported there were challenges also selected product characteristics that they found challenging. Multiple answers were allowed. Flavor, texture, and dough handling properties were the top three challenges. Though flavor was the fifth most selected satisfactory characteristic (**Figure 4.3A**), eight out of the 10 respondents selected flavor as a challenge (**Figure 4.3B**). It was interesting that texture was the most selected satisfactory characteristic, but six out of the 10 respondents selected it as a challenge. Texture is a blanket word that encompasses various attributes (e.g. firmness, brittleness, resilience) depending on product types (Bourré et al., 2019; Paladugula et al., 2021; Ramírez-Jiménez et al., 2018). Thus, individual research would be needed to address challenges regarding texture of specific products. Dough handling properties are also a recognized challenge due to lack of gluten in pulse flours (Lascialfari et al., 2019; Marti & Pagani, 2013; Melini et al., 2017). The results obtained from the end-users of pulse flours supported previous findings and highlighted the importance of flavor and texture.

Variability

The 21 current, previous, or considered users were asked questions about variability in pulse flour quality by batch or supplier, universal specifications of pulse flour or the lack thereof, and gluten and lectin concerns (**Figure 4.4**). Three respondents agreed that pulse flour quality (sometimes) varies from batch to batch, seven disagreed, and 10 did not know. One respondent who selected "Sometimes" commented that the quality is "harvest dependent". In contrast, for variation by suppliers, 11 users agreed that pulse flour quality (sometimes) varies from supplier to supplier. This seems to reflect the current situation in the pulse milling industry where there is no universal specifications for pulse flour products, unlike wheat flour, which results in pulse flour products with varying characteristics milled by the supplier's own method (Thakur et al., 2019).

Interestingly, however, not everyone agreed that the variation in pulse flour quality affects their product consistency or that having universal specifications is a critical factor for them. Two respondents who indicated that the variation in pulse flours "sometimes" makes it difficult to produce food products with consistent quality commented that it depends on varieties, growing locations, and weather conditions. These comments showed that some users are aware of the effect of pulse varieties and seasonal conditions on the flour quality and thus the food product quality made with it. Regarding universal specifications, one who selected "Sometimes" commented that "Depends on the level of incorporation of the pulse flour in the formulas", suggesting that specifications may not matter as much if only a small percentage of pulse flour was incorporated in other flours. Fifteen respondents indicated that gluten contamination is (sometimes) a concern in pulse flours, but only five indicated that lectins in pulse flours are a concern, similarly to the bean flour impression results (**Table 4.7**).

Supply and logistics

Extrinsic factors such as cost, production, and influence of the COVID-19 pandemic on pulse flour supply were assessed by the 21 users. Twelve said there is a supplier that provides them with their pulse flour at a reasonable cost (**Figure 4.5A**). It was intriguing that seven said freight cost is a critical factor, but five said it is not. Eight to ten participants either did not know or did not answer to these questions about supply and logistics. More than half (11) participants said there is enough supply of pulse flour for their production scale, while none said there is not enough supply. These results indicated that pulse flour supply or cost may not be the main barrier to pulse flour use for food manufacturers. The pandemic of COVID-19 saw a surge in sales of pulse products

(Domonoske & Schneider, 2020), but only three said pulse flour was less available while five said the availability is higher or the same (**Figure 4.5B**).

Limitations of the study

Response rate was low for this survey: the response rate was 0.85% (88/10,306) via email. The small number of responses violated the assumptions of chi-square test for proportion comparison among subgroups. The low response rate may be due to the industry's hesitation in answering questions about their production and research activities. Industry professionals are a difficult population to obtain information from even though it was clearly explained in the invitation email and survey front page that this survey is for academic research purposes and that no identifiable information will be published. An invitee selected "Not comfortable answering questions about the company activities" as a reason not to participate, and another invitee to the preliminary interview stated that they cannot participate because the questions are [asking information that is] mostly confidential for them. A similar observation was made in a pulse products innovation study; only two firms filed a patent on their new method of pulse processing (which will be published), and all others chose to have it as trade secret (Lascialfari et al., 2019). The self-administered form of survey may have contained questions or choices that participants did not understand. The percentage of participants in our target job tiles were low: only 40% (30/75) of the respondents were in R&D or production/chef, and others were in management/owner, regulation/QA/QC, and Sales/Marketing. Words such as flavor, texture, and dough handling properties could mean various attributes depending on products; thus, identifying specific attributes applicable to specific products would narrow down the food industry's needs for pulse flour improvement.

Conclusions

An online survey was administered with food industry professionals whose firms produced food products from regular wheat and/or gluten-free flours. Valid answers were collected from eight current, four previous, nine considered, and 54 non-users of pulse flours: 75 in total. Yeast breads were the most selected product type for which they are using, used, or would be interested in using pulse flours. Chickpea and pea were the top two pulse flours selected. Scientific associations, both industry and academic, and open-access journals were likely to reach industry personnel who look for technical information. Regarding bean flour in particular, increasing publicity and improving flavor and functionality may be the key for bean flour improvement for flour purposes. Lectin and gluten contaminations were less selected as a concern. Market demand may be more important for continued use of pulse flours than intrinsic characteristics. According to the 21 CPC users, texture, appearance, and uniformity were the three most selected satisfactory characteristics of products using their pulse flour. Ten stated that there were challenges, but five stated there were not. Flavor, texture, and dough handling properties were the top three most selected challenges. More than half of users agreed that pulse flour quality (sometimes) varies from supplier to supplier; however, opinions were mixed about universal specifications for pulse flours and the effect of the variation on product consistency. Extrinsic factors such as pulse flour supply or cost may not be a main problem to pulse flour users. In summary, flavor, texture, dough handling properties, and market demand were important factors in pulse flour use to food manufactures who participated in this survey. Individual characteristics that need improvements for specific pulse flour and product type should be studied in future research.

Acknowledgements

The authors would like to thank Dr. Tammy Long and her lab members in the Department of Plant Biology at Michigan State University for their advice on interviewing techniques. The authors also would like to express gratitude to Drs. Maurice Bennink, James Kelly, and Mark Uebersax at Michigan State University, and Mr. Joe Cramer at Michigan Bean Commission for their valuable advice on survey development.

APPENDICES

APPENDIX A:

CHAPTER 4 TABLES AND FIGURES

Table 4.1 SIC codes of food manufacturing firms targeted in this study.

SIC code	Product category
2045	Prepared Flour Mixes & Doughs
2051	Bread & Other Bakery Products
2052	Cookies & Crackers
2053	Frozen Bakery Products
2096	Potato Chips Corn Chips/similar Snacks
2098	Macaroni Spaghetti Vermicelli & Noodles

Table 4.2 Primary job roles of the respondents by user type.

	Times selected	Current	Previous	Considered	Never
	Count	---------- % ----------			
Management/Owner	29	25	25	22	44
R&D	18	50	25	67	13
Production/Chef	12	13	25	0	19
Regulation/QA/QC	9	0	25	0	15
Multiple roles	4	0	0	11	6
Sales/Marketing	3	13	0	0	4
Total	75	100	100	100	100

Table 4.3 Participants' currently using, previously used, considered using, or would be interested in using type of pulse flours by user type.

	Times selected	Current	Previous	Considered	Never
	Count	---------- % ----------			
Do not know which is good	37	0	0	0	70
Chickpea	13	38	0	56	9
Pea	13	63	50	33	6
Not interested	7	0	0	0	13
Bean	2	0	25	11	0
Faba	1	0	25	0	0
Lentil	1	0	0	0	2
Total	74	100	100	100	100

Table 4.4 Product type that the participants are currently using, previously used, considered using, or would be interested in using pulse flours for.

	Times selected	Current	Previous	Considered	Never
	Count	---------- % ----------			
Yeast breads, buns, and rolls	21	25	50	22	28
Cookies, bars, and crackers	11	0	0	11	19
Not interested	11	0	0	0	20
Quick breads such as muffins	9	25	25	0	11
Snack foods such as chips, puffs	6	13	25	33	2
Other	4	0	0	0	7
Pastries and pies	4	0	0	0	7
Plant-based meat alternatives	4	25	0	22	0
Cakes	3	0	0	11	4
Thickening agents for soups, gravies	2	13	0	0	2
Total	75	100	100	100	100

Table 4.5 Pulse flour type and product type that participants are using, previously used, considered using, or interested in using. A total 14 of "Not interested" or no answers were removed.

	Do not know which is good	Chickpea	Pea	Bean	Faba	Lentil	Total
Yeast breads, buns, and rolls	9	4	4	0	1	1	19
Cookies, bars, and crackers	9	0	2	0	0	0	11
Quick breads such as muffins	4	3	2	0	0	0	9
Snack foods such as chips, puffs	1	2	1	2	0	0	6
Other	2	1	1	0	0	0	4
Plant-based meat alternatives	0	2	2	0	0	0	4
Cakes	2	1	0	0	0	0	3
Pastries and pies	3	0	0	0	0	0	3
Thickening agents for soups, gravies	1	0	1	0	0	0	2
Total	31	13	13	2	1	1	61

Table 4.6 The number of participants who had (Yes) or had never (No) looked for technical information on pulse flour

	Times selected	Current	Previous	Considered	Never
	Count	---------- % ----------			
Yes	23	75	100	78	11
No	52	25	0	22	89
Total	75	100	100	100	100

Table 4.7 Impression on bean flour by current, previous, considered, and non-users of pulse flours.

	Times selected	Current	Previous	Considered	Never
	Count	---------- % ----------			
Do not know about bean flour	37	25	25	33	57
Flavor is a challenge	22	63	0	67	20
Functionality is a challenge	21	38	25	44	24
Market demand for bean flour is low	13	38	25	11	15
More expensive than other GF flours	10	25	25	0	13
Do not know how to use bean flour	9	13	0	22	11
Gluten contamination is a concern	5	13	0	11	6
Lack of specification for bean flour is a challenge	3	0	0	0	6
Not available from a local source	3	0	25	0	4
Other	3	0	0	0	6
Lectins in bean flour are a concern	1	0	25	0	0

Table 4.8 The combinations of product and pulse flour that the 21 current, previous, or considered users selected. Counts more than one are shown in parentheses.

Count	Product	Pulse flour type
6	Yeast breads, buns, and rolls	Pea (3), Chickpea (2), Faba
5	Snack foods such as chips, puffs	Bean (2), Chickpea (2), Pea
4	Plant-based meat alternatives	Pea (2), Chickpea, Other
3	Quick breads such as muffins	Pea (2), Chickpea
1	Cakes	Chickpea
1	Cookies, bars, and crackers	Pea
1	Thickening agents for soups, gravies	Pea

Table 4.9 Reasons for currently using pulse flours selected by eight current users.

Reason for using	Count
Driven by marketing and trends	5
Protein content	3
Functional Characteristics of the pulse flour	3
Environmental sustainability	2
Gluten-free attributes	2
Health benefits to consumers	1
Improved protein quality	1
Other	1

Table 4.10 Perceived presence or absence of challenges with pulse flours by user type.

		Current	Previous	Considered	Total
		---------- Count ----------			
Were there challenges?	Yes	3	2	5	10
	No	2	2	1	5
	Do not know	3	0	0	3
	Total	8	4	6	18

Table 4.11 Combination of product type and pulse flour type by the presence or absence of challenges in production. Counts more than one are shown in parentheses.

Challenges	Product type	Pulse flour type
Yes, there is (are) challenge(s).		
	Yeast breads, buns, and rolls	Chickpea (2), Pea (2), Faba
	Snack foods such as chips, puffs	Bean, Chickpea
	Plant-based meat alternatives	Chickpea, Pea
	Cookies, bars, and crackers	Pea
No, there was no challenge.		
	Quick breads such as muffins	Chickpea, Pea
	Snack foods such as chips, puffs	Bean, Chickpea
	Thickening agents for soups, gravies	Pea

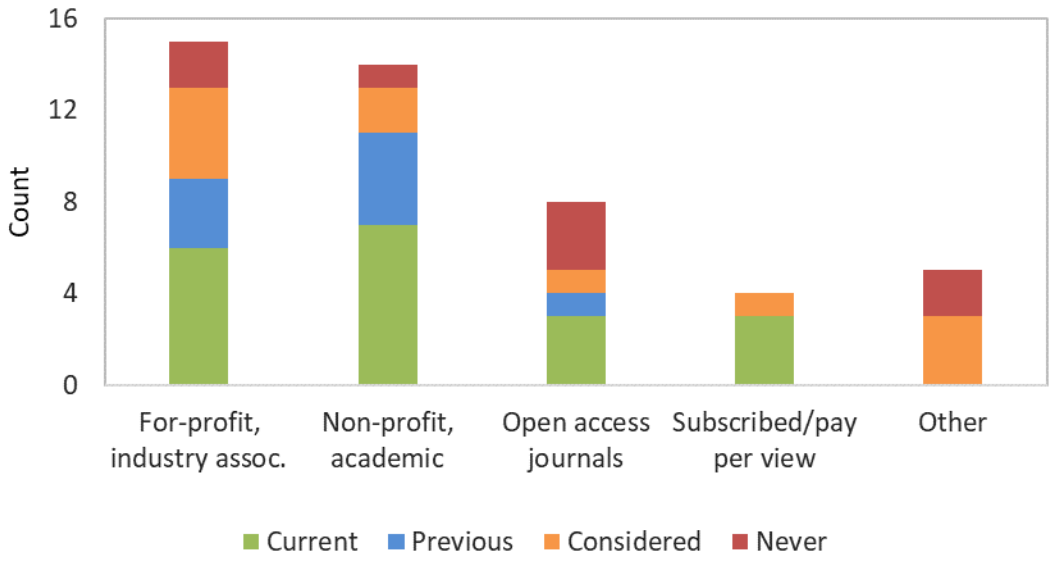


Figure 4.1 Source of information consulted by those who have looked for pulse flour technical information. For-profit, industry assoc. includes For-profit (industry) researchers and industry bodies (e.g., Institute of Food Technologists). Non-profit, academic includes non-profit organizations (e.g., American Pulse Association), and academic researchers and textbooks.

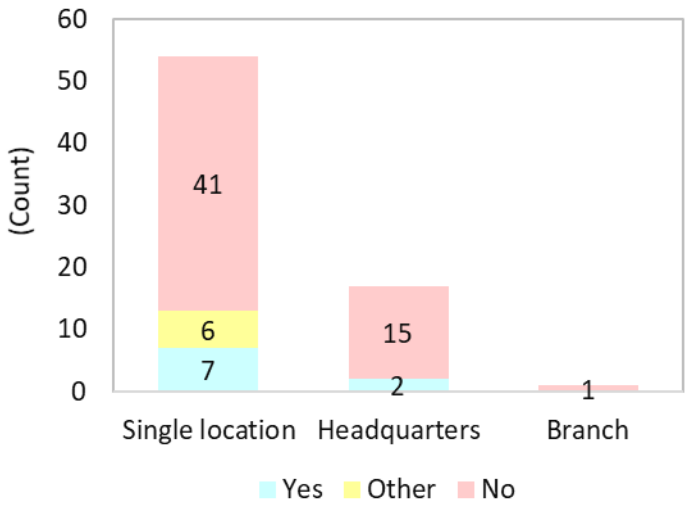
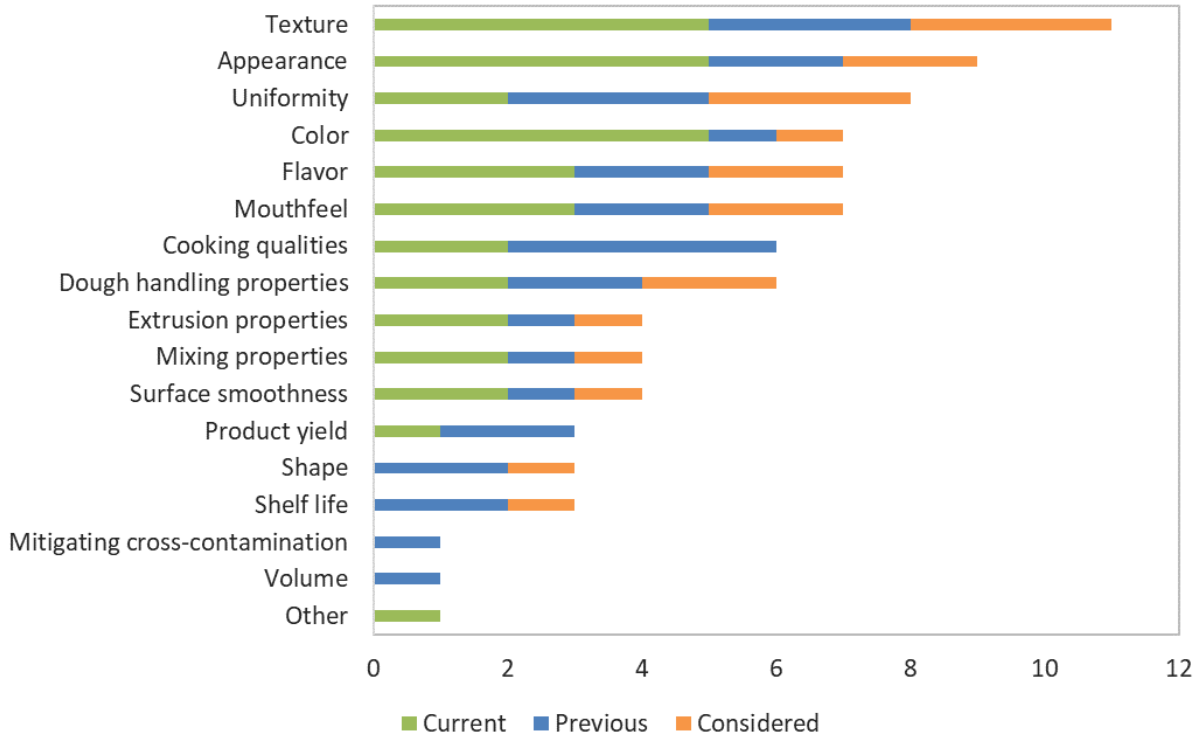


Figure 4.2 Business type and willingness to provide opinion for improving pulses for flour purposes.

A



B

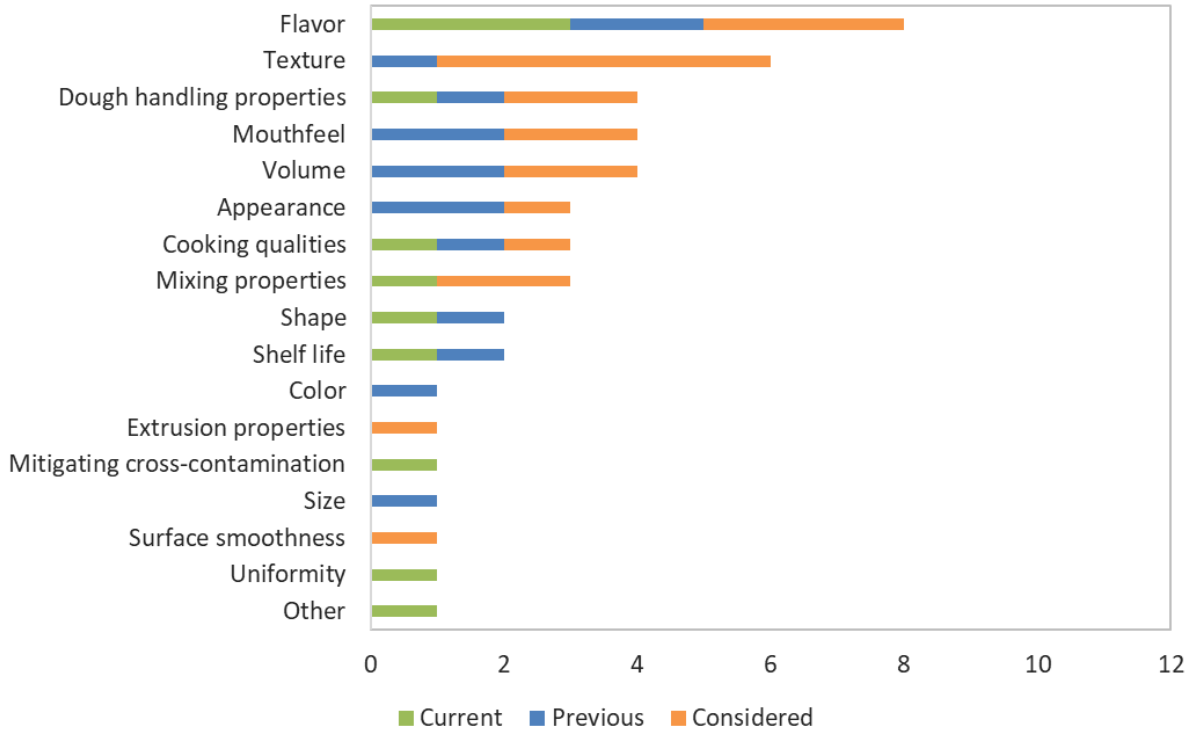


Figure 4.3 Satisfactory (**A**) and challenging (**B**) product qualities produced using the pulse flours that CPC users selected.

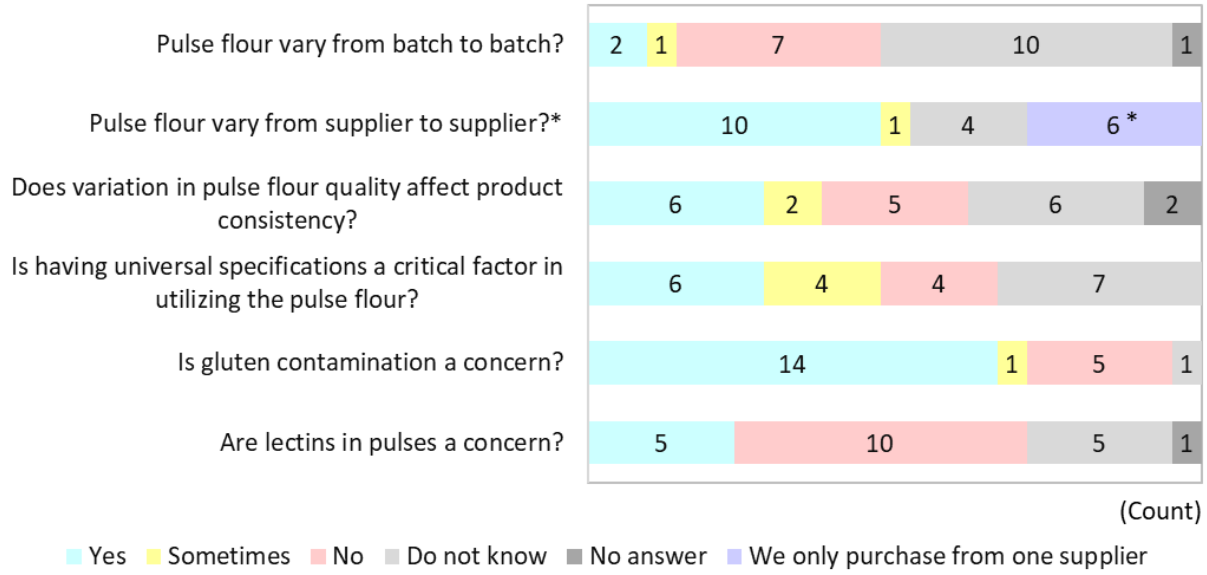
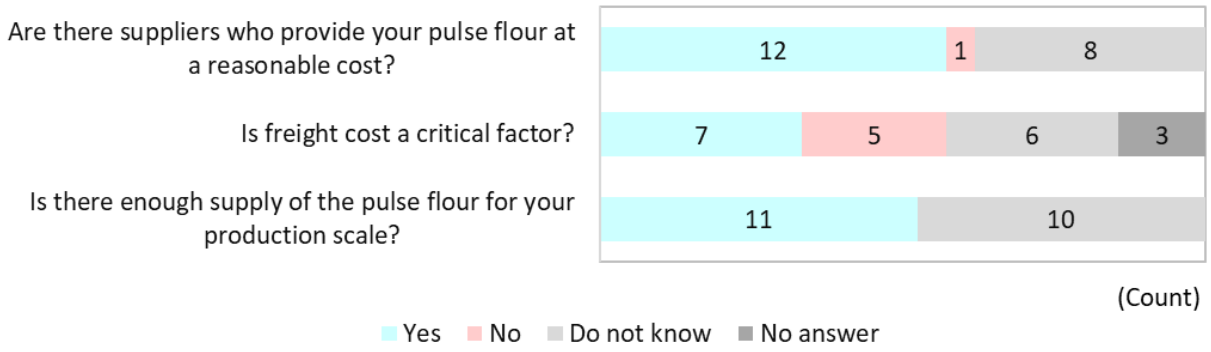


Figure 4.4 CPC users' opinions on pulse flour quality variability, universal specification, gluten contamination, and lectins. *The question about variation by suppliers had an additional choice: "We only purchase from one supplier".

A



B

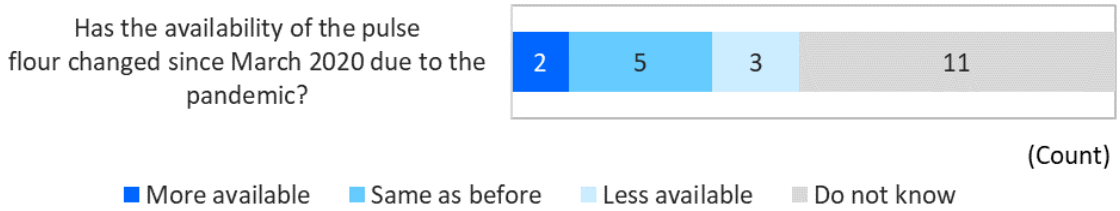


Figure 4.5 CPC users' opinions on the supply and logistics of pulse flours (**A**) and availability of pulse flours during the pandemic of COVID-19 (**B**).

APPENDIX B:

CHAPTER 4 SUPPLEMENTAL TABLES

Table S4.12 Job titles targeted in the second distribution of the survey.

Business Development
CEO
Chief Marketing Officer
Director
Engineering/Technical
Executive
Executive Director
Executive Officer
Executive Vice
President
General Manager
International
Manager
Manufacturing
Manufacturing
Executive
Marketing
Marketing Executive
Operations
Operations Executive
Owner
President
Principal
Purchasing Executive
Regional Manager
Sales
Sales Executive
Senior Vice President
Site Manager
Vice President

REFERENCES

REFERENCES

Abu-Ghannam, N., & Gowen, A. (2011). Pulse-based food products. In B. K. Tiwari, A. Gowen, & B. McKenna (Eds.), *Pulse Foods: Processing, Quality and Nutraceutical Applications* (pp. 249–282). Elsevier Inc. https://doi.org/10.1016/B978-0-1238-2018-1.00007-0

Alani, S. R., Zabik, M. E., & Uebersax, M. A. (1989). Dry roasted pinto bean (*Phaseolus vulgaris*) flour in quick breads. *Cereal Chemistry*, *66*, 348–349. https://www.aaccnet.org/publications/cc/backissues/1989/Documents/66_348.pdf

American Pulse Association. (2020). Pulse flours amid shifting consumer needs. *The Pulse Mill Newsletter*, *7*(9), 1–4.

Bourré, L., Frohlich, P., Young, G., Borsuk, Y., Sopiwnyk, E., Sarkar, A., Nickerson, M. T., Ai, Y., Dyck, A., & Malcolmson, L. (2019). Influence of particle size on flour and baking properties of yellow pea, navy bean, and red lentil flours. *Cereal Chemistry*, *96*(4), 655–667. https://doi.org/10.1002/cche.10161

Bresciani, A., & Marti, A. (2019). Using pulses in baked products: Lights, shadows, and potential solutions. *Foods*, *8*(10), 451. https://doi.org/10.3390/foods8100451

Cargill. (2019). *PURIS™ Pea Protein*. https://www.cargill.com/food-bev/na/pea-protein

Dabija, A., Codina, G. G., & Fradinho, P. (2017). Effect of yellow pea flour addition on wheat flour dough and bread quality. *Romanian Biotechnological Letters*, *22*(5), 12888–12897. https://e-repository.org/rbl/vol.22/iss.5/5.pdf

Domonoske, C., & Schneider, A. (2020). *Here's what's been flying off store shelves*. NPR Special Series The Coronavirus Crisis. https://www.npr.org/2020/03/16/816404689/spiking-demand-for-sanitizer-canned-goods-leaves-stores-struggling-to-keep-up

Dryer, S. B., Phillips, S. G., Powell, T. S., Uebersax, M. A., & Zabik, M. E. (1982). Dry roasted navy bean flour incorportaion in a quick bread. *Cereal Chemistry*, *59*(4), 319–320. http://www.aaccnet.org/publications/cc/backissues/1982/Documents/Chem59_319.pdf

FAOSTAT. (2021). *No Title*. Food and Agriculture Organization of the United Nations. http://www.fao.org/faostat/en/#data/QC

Farooq, Z., & Boye, J. I. (2011). Novel food and industrial applications of pulse flours and fractions. In B. K. Tiwari, A. Gowen, & B. McKenna (Eds.), *Pulse Foods* (pp. 283–323). Elsevier. https://doi.org/10.1016/B978-0-12-382018-1.00011-3

Gallegos-Infante, J. A., Rocha-Guzman, N. E., Gonzalez-Laredo, R. F., Ochoa-Martinez, L. A., Corzo, N., Bello-Perez, L. A., Medina-Torres, L., & Peralta-Alvarez, L. E. (2010). Quality of spaghetti pasta containing Mexican common bean flour (*Phaseolus vulgaris* L.). *Food Chemistry*, *119*(4), 1544–1549. https://doi.org/10.1016/j.foodchem.2009.09.040

Hall, C., Hillen, C., & Garden Robinson, J. (2017). Composition, nutritional value, and health benefits of pulses. *Cereal Chemistry Journal*, *94*(1), 11–31. https://doi.org/10.1094/CCHEM-03-16-0069-FI

Han, J. (Jay), Janz, J. A. M., & Gerlat, M. (2010). Development of gluten-free cracker snacks using pulse flours and fractions. *Food Research International*, *43*(2), 627–633. https://doi.org/10.1016/j.foodres.2009.07.015

Hooper, S. D., Glahn, R. P., & Cichy, K. A. (2019). Single varietal dry bean (*Phaseolus vulgaris* L.) pastas: Nutritional profile and consumer acceptability. *Plant Foods for Human Nutrition*, *74*(3), 342–349. https://doi.org/10.1007/s11130-019-00732-y

Lascialfari, M., Magrini, M.-B., & Triboulet, P. (2019). The drivers of product innovations in pulse-based foods: insights from case studies in France, Italy and USA. *Journal of Innovation Economics*, *28*(1), 111143. https://doi.org/10.3917/jie.028.0111

Malcolmson, L., & Han, J. (2019). Pulse processing and utilization of pulse ingredients in foods. In W. J. Dahl (Ed.), *Health Benefits of Pulses* (pp. 129–149). Springer International Publishing. https://doi.org/10.1007/978-3-030-12763-3_9

Marti, A., & Pagani, M. A. (2013). What can play the role of gluten in gluten free pasta? *Trends in Food Science and Technology*, *31*(1), 63–71. https://doi.org/10.1016/j.tifs.2013.03.001

Martínez, M. M. (2015). *Applications of the rapid visco analyser (RVA) in the food industry: a broader view*. Perten Articles.

Maskus, H. (2010). Pulse Processing, Functionality and Application. In *University of Winnipeg* (pp. 1–146).

Melini, F., Melini, V., Luziatelli, F., & Ruzzi, M. (2017). Current and forward-looking approaches to technological and nutritional improvements of gluten-free bread with legume flours: A critical review. *Comprehensive Reviews in Food Science and Food Safety*, *16*(5), 1101–1122. https://doi.org/10.1111/1541-4337.12279

Mitchell, D. C., Lawrence, F. R., Hartman, T. J., & Curran, J. M. (2009). Consumption of Dry Beans, Peas, and Lentils Could Improve Diet Quality in the US Population. *Journal of the American Dietetic Association*, *109*(5), 909–913. https://doi.org/10.1016/j.jada.2009.02.029

Mora-Avilés, A., Lemus-Flores, B., Miranda-López, R., Hernández-López, D., Pons-Hernández, J. L., Acosta-Gallegos, J. A., & Guzmán-Maldonado, S. H. (2007). Effects of common bean enrichment on nutritional quality of tortillas produced from nixtamalized regular and quality protein maize flours. *Journal of the Science of Food and Agriculture*, *87*(5), 880–886. https://doi.org/10.1002/jsfa.2801

Mudryj, A. N., Yu, N., & Aukema, H. M. (2014). Nutritional and health benefits of pulses. *Applied Physiology, Nutrition, and Metabolism*, *39*(11), 1197–1204. https://doi.org/10.1139/apnm-2013-0557

Paladugula, M. P., Smith, B., Morris, C. F., & Kiszonas, A. (2021). Incorporation of yellow pea flour into white pan bread. *Cereal Chemistry*, cche.10441. https://doi.org/10.1002/cche.10441

Palmer, S., Winham, D., Oberhauser, A., & Litchfield, R. (2018). Socio-ecological barriers to dry grain pulse consumption among low-income women: A mixed methods approach. *Nutrients*, *10*(8), 1108. https://doi.org/10.3390/nu10081108

Parr, B., & Lucier, G. (2020). Vegetables and pulses yearbook data. *USDA-ERS Situation and Outlook Report, #89011*.

Ramírez-Jiménez, A. K., Gaytán-Martínez, M., Morales-Sánchez, E., & Loarca-Piña, G. (2018). Functional properties and sensory value of snack bars added with common bean flour as a source of bioactive compounds. *LWT*, *89*, 674–680. https://doi.org/10.1016/j.lwt.2017.11.043

Roland, W. S. U., Pouvreau, L., Curran, J., van de Velde, F., & de Kok, P. M. T. (2017). Flavor aspects of pulse ingredients. *Cereal Chemistry Journal*, *94*(1), 58–65. https://doi.org/10.1094/CCHEM-06-16-0161-FI

Simons, C. W., & Hall, C. (2018). Consumer acceptability of gluten-free cookies containing raw cooked and germinated pinto bean flours. *Food Science & Nutrition*, *6*(1), 77–84. https://doi.org/10.1002/fsn3.531

Sozer, N., Holopainen-Mantila, U., & Poutanen, K. (2017). Traditional and new food uses of pulses. *Cereal Chemistry Journal*, *94*(1), 66–73. https://doi.org/10.1094/CCHEM-04-16-0082-FI

Sparvoli, F., Laureati, M., Pilu, R., Pagliarini, E., Toschi, I., Giuberti, G., Fortunati, P., Daminati, M. G., Cominelli, E., & Bollini, R. (2016). Exploitation of common bean flours with low antinutrient content for making nutritionally enhanced biscuits. *Frontiers in Plant Science*, *7*, 1–14. https://doi.org/10.3389/fpls.2016.00928

Szczygiel, E. J., Harte, J. B., Strasburg, G. M., & Cho, S. (2017). Consumer acceptance and aroma characterization of navy bean (*Phaseolus vulgaris*) powders prepared by extrusion and conventional processing methods. *Journal of the Science of Food and Agriculture*, *97*(12), 4142–4150. https://doi.org/10.1002/jsfa.8284

Teodoro, M. (2017). Flour power: Ingredient trends in flours and alternative flours. *Ingredient Insight, Mintel*, 1–29.

Thakur, S., Scanlon, M. G., Tyler, R. T., Milani, A., & Paliwal, J. (2019). Pulse Flour Characteristics from a Wheat Flour Miller's Perspective: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, *18*(3), 775–797. https://doi.org/10.1111/1541-4337.12413

US Department of Agriculture, & US Department of Health and Human Services. (2015). *2015 – 2020 Dietary Guidelines for Americans. 8th Edition*. https://doi.org/10.1097/NT.0b013e31826c50af

Winham, D., Tisue, M., Palmer, S., Cichy, K., & Shelley, M. (2019). Dry bean preferences and attitudes among Midwest Hispanic and non-Hispanic white women. *Nutrients, 11*(1), 178. https://doi.org/10.3390/nu11010178

Wood, J. A., & Malcolmson, L. J. (2011). Pulse milling technologies. In B. K. Tiwari, A. Gowen, & B. McKenna (Eds.), *Pulse Foods: Processing, Quality and Nutraceutical Applications* (pp. 193–221). Academic Press. https://doi.org/10.1016/B978-0-1238-2018-1.00007-0

Zarrouki, K. (2017). New products containing pulse ingredients in North America. *Global Analysis Report, The Minister of Agriculture and Agri-Food, Canada*, 1–15. https://www.agr.gc.ca/eng/international-trade/market-intelligence/reports/commodity-innovation-series-new-products-containing-pulse-ingredients-in-north-america/?id=1499950769666#k