Theor Appl Genet (2015) 128:1555–1567
DOI 10.1007/s00122-015-2531-z

ORIGINAL PAPER

Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.)

Karen A. Cichy[1,2] · Jason A. Wiesinger[2] · Fernando A. Mendoza[2,3]

Received: 18 March 2015 / Accepted: 27 April 2015 / Published online: 24 May 2015

Abstract

Key message **Fivefold diversity for cooking time found in a panel of 206 *Phaseolus vulgaris* accessions. Fastest accession cooks nearly 20 min faster than average. SNPs associated with cooking time on Pv02, 03, and 06.**

Abstract Dry beans (*Phaseolus vulgaris* L.) are a nutrient dense food and a dietary staple in parts of Africa and Latin America. One of the major factors that limits greater utilization of beans is their long cooking times compared to other foods. Cooking time is an important trait with implications for gender equity, nutritional value of diets, and energy utilization. Very little is known about the genetic diversity and genomic regions involved in determining cooking time. The objective of this research was to assess cooking time on a panel of 206 *P. vulgaris* accessions, use genome- wide association analysis (GWAS) to identify genomic regions influencing this trait, and to test the ability to predict cooking time by raw seed characteristics. In this study 5.5-fold variation for cooking time was found and five bean accessions were identified which cook in less than 27 min across 2 years, where the average cooking time was 37 min. One accession, ADP0367 cooked nearly 20 min faster than average. Four of these five accessions showed close phylogenetic relationship based on a NJ tree developed with ~5000 SNP markers, suggesting a potentially similar underlying genetic mechanism. GWAS revealed regions on chromosomes Pv02, Pv03, and Pv06 associated with cooking time. Vis/NIR scanning of raw seed explained 68 % of the phenotypic variation for cooking time, suggesting with additional experimentation, it may be possible to use this spectroscopy method to non-destructively identify fast cooking lines as part of a breeding program.

Introduction

Food choices by humans are based on many factors including cost, convenience, accessibility, taste, and nutrition (Furst et al. 1996). Trade-offs are often made among these factors, especially related to nutritional value (Guthrie et al. 2002). In the U.S. convenience is very important since the average household only spends 60 min. per day in meal preparation, or about 20 min. per meal (Smith et al. 2013). There is a need and a demand for nutritious food options that do not require long preparation times. While in the U.S. and other developed countries, long cooking times have been circumvented by utilization of processed foods, in developing counties that is not always an option and meal preparation is more labor intensive. Globally about 40 % of the population uses firewood and charcoal as their primary fuel used for cooking; in Sub Saharan Africa it is as high as 76 % (Felix and Gheewala 2011; Maes and Verbist 2012). Gathering of firewood is a time-consuming activity, which is typically the responsibility of women and children in African households (Menendez and Curt 2013). Utilization

Communicated by G. J. Bryan.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2531-z) contains supplementary material, which is available to authorized users.

✉ Karen A. Cichy
karen.cichy@ars.usda.gov

[1] Sugarbeet and Bean Research Unit, USDA-ARS, 1066 Bogue St., Room 494, East Lansing, MI 48824, USA

[2] Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA

[3] Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

of nutritious, inexpensive, fast cooking foods could potentially have a positive impact for women in these regions.

Dry beans (*Phaseolus vulgaris* L.) are a food for which improvements in cooking time would be especially valuable. Beans are a dietary staple in parts of Africa and Latin America where moderate to high consumption (16–34 kg per capita annually) is common (Akibode and Maredia 2011). While per capita dry bean consumption rate is low (2.8 kg annually) in the U.S., they are still one of the top ten most consumed vegetables/pulses (Drewnowski and Rehm 2013; Siddiq et al. 2012). Dry beans are a rich source of protein, dietary fiber, iron, magnesium, and folate (Bressani 1983; Winham et al. 2008). Despite their high nutritional value, beans require long cooking times to become palatable. It takes 7–11 kg of fuel wood to cook one kg of beans, in contrast to one kg of maize flour, which requires less than one kg of fuel wood to cook (Adkins et al. 2010). Decreasing the cooking times of dry beans would be especially important in areas where beans are consumed as a primary source of protein.

Cooking of beans is a hydrothermal process where chemical and physical changes occur within the seed, including starch swelling and gelatinization, protein denaturation, and partial solubilization of some cell wall polysaccharides (Rockland and Jones 1974; Vindiola et al. 1986). The cooking time of beans is influenced by the storage environment. Freshly harvested beans cook 2–4 times faster than beans stored for 6 months (Coelho et al. 2007). Storing beans for long periods of time, especially in high temperatures and high humidity, greatly increases the cooking time, often referred to as the 'hard to cook' phenomenon (Liu 1995). A bean that cooked in 1 h just after harvest can take 2–3 h to cook as it ages (Hernández-Unzón and Ortega-Delgado 1989). There are numerous physiological changes that take place during storage that influence cooking time. These include the breakdown of phytic acid by the enzyme phytase, which frees divalent cations (especially calcium) to cross link with cell wall polysaccharides (Jones and Boulter 1983; Mattson et al. 1950). In addition, polyphenols become less extractable in hard to cook beans (Stanley et al. 1990). Previous work shows long cooking times that result from the 'hard to cook' phenomenon reduces the protein nutritional quality of dry beans (Tuan and Phillips 1992).

The genetic variability for cooking time is less understood than the environmental influences on cooking time. In the evaluation of a small number of bean accessions, twofold variation in cooking time has been noted (Elia et al. 1997). Very few studies have been conducted on the genetic control of cooking time in beans. These studies indicate this trait is controlled by a small number of genes and is highly heritable with narrow sense heritability values between 0.74 and 0.90 (Elia et al. 1997; Jacinto-Hernandez et al. 2003). The accurate determination of cooking time is laborious and a limited number of samples can be processed in a day. This is a trait for which faster evaluation methods and more information on the genetic control are needed by breeders. The objective of this study was to evaluate diversity of cooking time on a panel of over 200 dry bean accessions (mostly of Andean origin) and to use genome-wide association analysis (GWAS) to identify genomic regions involved in determining cooking time. In addition, the ability to predict cooking time with Vis/NIR spectroscopy on raw seed was evaluated.

Materials and methods

Germplasm

A subset of 206 *Phaseolus vulgaris* accessions of the Andean Diversity Panel (ADP) was used for this experiment. The genetic composition of the entire ADP is described in Cichy et al. (2015). The ADP subset includes accessions from Africa (93), Europe (6), Asia (3), Central America (7), Caribbean (18), North America (68), and South America (11).

Field design

These 206 ADP accessions were grown at the Michigan State University, Montcalm Research Farm near Entrican, MI, in 2012 and 2013. The soil type is Eutric Glossoboralfs (coarse-loamy, mixed) and Alfic Fragiorthods (coarse-loamy, mixed, frigid). The materials were planted in a Randomized Complete Block Design with two replications. Plots consisted of two rows 4.75 m long with 0.5 m spacing between rows. Rainfall was supplemented with overhead irrigation as needed. No fertilizer was applied to the plots and recommended practices were followed for weed and insect control. Seed was harvested upon maturity by hand pulling plants and threshing with a plot Hege 140 plot harvester (Wintersteiger, Utah, USA). After harvest seed was hand cleaned to remove debris and damaged seed.

There was a larger subset of 250 accessions planted, but the set was reduced to 206 for inclusion in the analysis as follows: 21 accessions were unadapted to Michigan growing conditions and seed was not consistently produced in both years, and 23 accessions were removed because the cooking time values were inconsistent across field replication and/or years and were identified as outliers with Z scores from Grubbs test of 2.5 or greater (Grubbs 1950). In addition data from 3 lines ADP 213, 433, and 442 were removed from water uptake because the coefficient of variation across field reps was greater than 25 %. These inconsistencies may have been due to the lack of adaptation to

Michigan growing conditions and, therefore, they were removed from the analyses.

Phenotypic evaluation

Visible and near-infrared reflectance (Vis/NIR) spectroscopy analysis

Freshly harvested seed from both field reps of each of the ADP accessions (2013 field season only) were scanned via Vis/NIR. A laboratory Vis/NIR spectrophotometer (Model 6500, Foss NIR Systems Inc., Silver Springs, MD, USA) was used to acquire reflectance spectra from intact dry beans in the range of 400–2498 nm collected at increments of 2 nm, yielding each spectrum of 1050 wavelengths. The Model 6500 is a pre-dispersive grating NIR instrument with a dual silicon and PbS detector system which allows full range measurement of visible and near infrared from 400 to 2500 nm. For scanning, the sample cell (Model Natural Product Sample Cell, Foss NIR Systems Inc., Silver Springs, MD, USA) was filled and smoothly compacted/leveled to provide a measurement stable against sample orientation. Then, multiple linear regression (MLR) models were evaluated to explain cooking time variability. For model testing, combinations of increased number of selected features were evaluated for improving the prediction accuracy of the regression model. The best subset of Vis/NIR wavelengths, after preprocessing by applying a first derivative to the data (that leads to the smallest prediction error), was found using the sequential forward selection (SFS) algorithm.

Water uptake and cooking time

A subsample of approximately 200 seeds was randomly selected and placed in a 6.5 × 14 cm envelop and stored at room temperature. To equilibrate moisture content before cooking, samples were held in hermetic storage over a saturated salt solution (63 % relative humidity) within a controlled atmosphere storage cabinet (Storage Control Systems, Inc. Sparta, MI, USA) prior to cooking. When the moisture content reached 10–14 %, as determined with a moisture meter (Moisture Check Plus, Deere and Company, Moline, IL), 30 seeds were selected per entry, weighed, and soaked in (1:4) w/w distilled water for 12 h. Seed were then drained, blotted dry, and weighed again to determine percent water uptake during soaking with the equation: [(seed weight after soak–seed weight before soak)/seed weight before soak] × 100. Cooking time was measured on 25 soaked seeds with a Mattson cooker (Customized Machining and Hydraulics Co., Winnipeg, Canada) in boiling distilled water. The Mattson Cooker consists of a plate with 25 wells for individual seeds and on top of each well rests a metal pin (Wang and Daun 2005). Cooking was recorded as the time it took for 80 % of the seeds to be completely pierced with an 85 g stainless steel rod with a 2-mm pin.

Data analysis

Statistical analysis of the phenotypic data was conducted with SAS version 9.2 of the SAS System for Windows (SAS Institute Inc. Cary, NC, USA). Normality was tested with the Shapiro–Wilk test (Shapiro and Wilk 1965). The GLM procedure was used for analysis of variance with the model including genotype (206 levels) year (2 levels) and rep (2 levels) and genotype × year. The CORR procedure was used to determine Pearson's correlation coefficients for each trait of interest.

Phylogenetic distance analysis was conducted using the method of Nei et al. (1983) with the set of 4935 SNP markers which remained from the original set after removing monomorphic markers. The analysis was conducted and a neighbor joining tree with 1000 bootstrapping iterations was developed using PowerMarker v3.25 (Liu and Muse 2005). The tree graphic was developed in Mega6 (Tamura et al. 2013).

Genome-wide association analysis

The 206 lines were genotyped with the Illumina (Illumina Inc., San Diego, CA, USA) BARCBean6K_3 SNP array of 5,398 SNP markers distributed across the 11 pairs of common bean chromosomes. The BARCBean6K_3 BeadChips were scanned with the Illumina BeadStation 500G. SNP calling was conducted with the genotyping module V2011.1 of GenomeStudio software (Illumina Inc., San Diego, CA, USA) as described in (Cichy et al. 2015). The entire set of 5398 SNP markers was filtered to remove monomorphic SNPs and those SNPs with a minor allele frequency of 2 % or less. After filtering, 3741 SNPs remained for GWAS analysis with a mixed linear model (MLM). Population structure was accounted for in the model with Principal Component Analysis (PCA) using a correlation matrix in the program TASSEL 4.0 (Bradbury et al. 2007) and two PCs were included, which explained 37.3 % of the variance (PC1 28.5 %, PC2 8.86 %). A kinship matrix (*K*) was also included in the association analysis to account for relatedness. The QQ plot was generated from the observed and expected LOD scores for each trait. Manhattan plot and QQ plot graphics were developed in qqman in *R* (Turner 2014).

The phenotypic data for cooking time and water uptake were both averaged across the 2012 and 2013 field seasons prior to GWAS. Cooking time data were transformed with the Box-Cox method (Box and Cox 1964). The following

MLM equation was used: $Y = X\alpha + P\beta + K\mu + e$, where Y is phenotype, X is SNP, p is the PCA matrix and both X and p represent fixed effects, K is the relative kinship matrix value, and e is for residual effects. The cutoff used for significant SNP markers for each trait was the False Discovery Rate (FDR) (Benjamini and Hochberg 1995) determined in using Bioconductor in R (Gentleman et al. 2004). In addition markers above a 2.5 LOD score were also considered as potential candidates. Candidate genes were identified based on proximity to significant SNPs using the *P. vulgaris* reference genome (Schmutz et al. 2014).

Results

Phenotypic diversity

A subset of 206 accessions of the ADP was evaluated for traits related to cooking time over two field seasons. The ADP includes 396+ *P. vulgaris* accessions from around the globe that are being used as part of a collaborative effort to develop improved cultivars of Andean beans especially targeted for use in Africa and the Americas (Cichy et al. 2015). Seed fresh weight as determined per 30 seed was normally distributed and ranged from 7.9 to 16.2 g. The average seed weight was 14.7 g in 2012 and was greater in 2013 at 16.2 g. Genotype, year, and genotype by year interactions were all significant for seed weight (Table 1). The 30 seeds used to measure seed weight were soaked in distilled water for 12 h. Percent water uptake during soaking ranged from 55 to 130 %. This trait was not normally distributed but this was largely due to a few outliers with very low water uptake. Overall the values were skewed toward higher percent water uptake (data not shown). Average water uptake was the same for both years at 106 %. Genotype was the only factor associated with water uptake in an analysis of variance (Table 1). After the 12-h soak, a set of 25 seed per field replicate were cooked with a Mattson cooker.

Within the 206 accessions 5.5-fold variability for cooking time was observed. In 2012, the cooking time averaged 36.4 min and in 2013 it was nearly 2 min longer on average at 38.3 min. The shortest cooking time was 16.5 min and the longest was 90 min, both observed in 2012 (Table 1). Cooking time was not normally distributed and was skewed toward longer times (data not shown). Significant genotype and year effects were identified in the analysis of variance, but genotype × year was not significant (Table 1). For further analyses water uptake and cooking time measurements were averaged across years.

Numerous seed types and colors are represented in the ADP, including major market classes in North America: light red kidney, dark red kidney, white kidney, and cranberry, as well as major market classes in Africa: red mottled, purple speckled, white and yellow. Accessions were grouped by market class to determine the importance of seed type in observed variability for cooking time. On average, white beans had the fastest cooking time at 30.2 min while the red/purple mottled had the longest cooking time at 41.1 min (Table 2). The ranges and standard deviations by market class were also evaluated. A box plot of the cooking time data by seed type reveals that within some market classes, especially cranberry and red mottled, there was greater variability for cooking time than for other market classes such as white and purple speckled where the variability among individual accessions was lower (Fig. 1).

Table 1 Mean, Median, and Range of seed fresh weight, water uptake during soaking, and cooking time of 206 *P. vulgaris* accessions of the Andean Diversity Panel grown in Entrican, MI in 2012 and 2013

	Min	Median	Mean	Max	CV %[a]	W[b]	Genotype	Year	G × Y[c]
Seed weight (g per 30 seed)									
2012	7.9	14.9	14.7	21.2	7.5	0.144	<0.0001	<0.0001	<0.0001
2013	7.9	16.3	16.2	24.0		0.588			
Water uptake after 12 h soak (%)									
2012	55	107	106	128	6.7	<0.0001	<0.0001	NS	NS
2013	61	106	106	130		<0.0001			
Cook time (min)									
2012	16.5	35.0	36.4	90	15.7	<0.0001	<0.0001	<0.0001	NS
2013	18.0	36.3	38.3	83.2		<0.0001			

p values associated normality tests (W) and with Type III sum of squares for genotype, year, and genotype × year interactions are presented

Values below 0.05 indicate non-normal distribution

[a] Coefficient of variation

[b] P value associated with Shapiro–Wilk normality tests

[c] Genotype × year interaction

Table 2 Mean cooking time of 206 Andean Diversity Panel accessions averaged across 2012 and 2013 evaluations and grouped based on seed types

Seed type	N[b]	Cook time (min)	
		Mean	St. Dev.
Cranberry	31	40.5	11.6
Dark red kidney	36	39.4	7.7
Light red kidney	30	39.3	6.9
Purple/red speckled	16	34.7	5.9
Red/purple mottled	22	41.1	14.0
Small red	8	37.3	8.6
White	15	30.2	4.8
Yellow	32	35.0	7.4
Other	16	32.8	3.6

[a] The *p* value of the analysis of variance was 0.0004

[b] Number of accessions

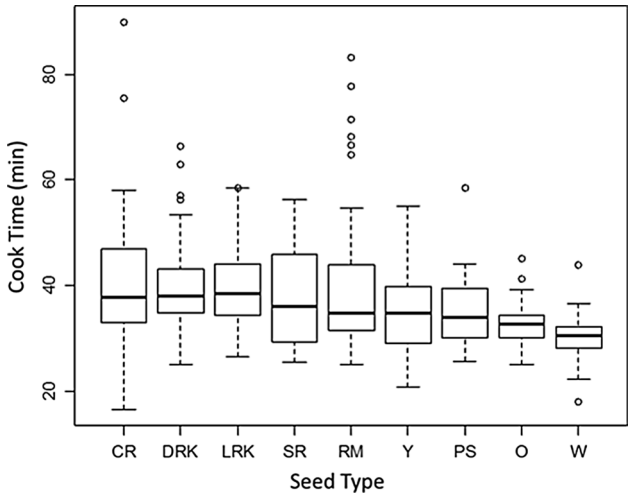

Fig. 1 *Boxplot* of cooking time for 206 Andean Diversity Panel accessions grouped by seed type. Mean cooking time for 2012 and 2013 are included as separate data points. Seed type abbreviations are as follows: *CR* cranberry, *DRK* dark red kidney, *LRK* light red kidney, *SR* small red, *RM* red mottled, *Y* yellow, *PS* purple speckled, *O* other, *W* white

Table 3 Mean cooking time of 206 Andean Diversity Panel accessions averaged across 2012 and 2013 evaluations and grouped based on region of origin

Origin	N[b]	Cook time (min)	
		Mean	St. Dev.
Africa	93	34.7	7.0
Asia	3	34.9	7.0
Caribbean	18	46.0	14.1
Cent. Am	7	37.0	7.4
Europe	6	35.8	8.0
N. Am	68	40.1	9.0
S. Am	11	31.6	5.1

[a] The *p* value of the analysis of variance was <0.0001

[b] Number of accessions

Table 4 Mean cooking time of 206 Andean Diversity Panel accessions averaged across 2012 and 2013 evaluations and grouped based on cultivation status

Cultivation	N[b]	Cook time (min)	
Status		Mean	St. Dev.
Landrace	108	35.8	7.5
Breeding Line	23	38.8	10.1
Variety	75	39.3	10.7

[a] The *p* value of the analysis of variance was 0.034

[b] Number of accessions

Table 5 Pearson correlation (*R*) and associated *p* values for cooking time with seed weight, water uptake, seed type, origin, and cultivation status

Variable	*R* with cook time	*p* values
Seed weight	0.100	0.043
Water uptake	−0.134	0.007
Seed type	0.285	<0.0001
Region of origin	0.281	<0.0001
Cultivation status	0.155	0.002

Origin is another factor that was considered in the analysis. This set of germplasm was gathered from diverse locations with the majority from Africa (45 %) and North America (33 %). Asia, Caribbean, Central America, and South America were also represented in the panel. The average cooking time of germplasm from the Caribbean was the highest at 46 min followed by North America at 40.1 min. The fastest cooking germplasm came from South America at 31.6 min followed by Africa and Asia at about 35 min (Table 3).

Cultivation status was a factor of interest since it points to potential social and market value. Accessions were classified as either landraces, breeding lines or released varieties. Landraces cooked on average 3–4 min faster than the breeding lines and varieties (Table 4). This suggests that in some cases farmers may select and maintain faster cooking germplasm, whereas this trait has not been a part of major bean breeding programs so their averages are slightly higher.

Pearson correlations revealed a weak but significant correlation between cooking time and seed weight such that larger seed took longer to cook. There was also a weak, negative correlation with water uptake, such that accessions

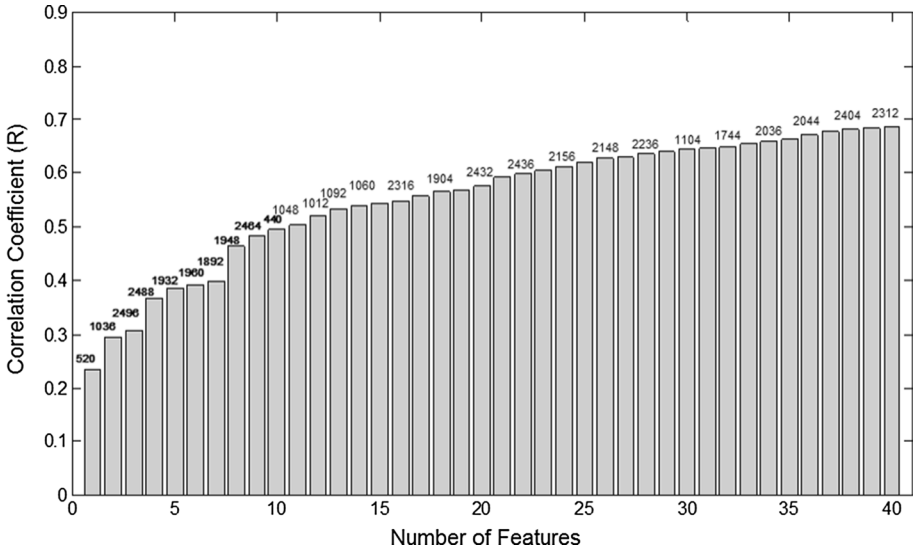

Fig. 2 Performance of multiple regression models using increased number of features or wavelengths for predicting cooking time in dry bean seeds. The best subset of Vis/NIR data after preprocessing was found using the sequential forward selection algorithm. Numbers on the top of the *bars* represent some of the selected wavelengths using sequential forward selection (SFS) algorithm

that took up less water during soaking took longer to cook. The highest correlations were with seed type and origin (Table 5). These correlations point to seed coat color playing a role in cooking time. These findings suggest that cooking time is a more important trait in some parts of the world than others. However, none of these correlations were large enough to be used as proxies in the measurement of cooking time.

Vis/NIR prediction

The possibility of a proxy for cooking time was explored further with this data set and Vis/NIR data was gathered on intact seed within 1 month after harvest. MLR models were tested based on the combination of increased number of Vis/NIR data. The best subset of 40 Vis/NIR features or wavelengths after applying first derivative was found using the sequential forward selection (SFS) algorithm. Figure 2 depicts the correlation coefficients (R) of the MLR models increasing the number of the best selected features in the model. As expected, better results were obtained by increasing the number of features in the predictive model. The R values improved from 0.235 (using only the best Vis/NIR band at 520 nm) to 0.690 when 40 features were considered in the prediction model. However, no significant improvements in predicting cooking time were observed using more than 37 features in the MLR model, giving an R value of 0.686. It should be also noted that among the best 40 Vis/NIR features selected by the SFS method, the wavelength at 520 nm was the first most important variable contributing to the model building. The spectral band at 520 nm was the only feature selected from the visible range and it is usually related to the anthocyanin content (Wang et al. 2013). The rest of the features were selected from the near-infrared region

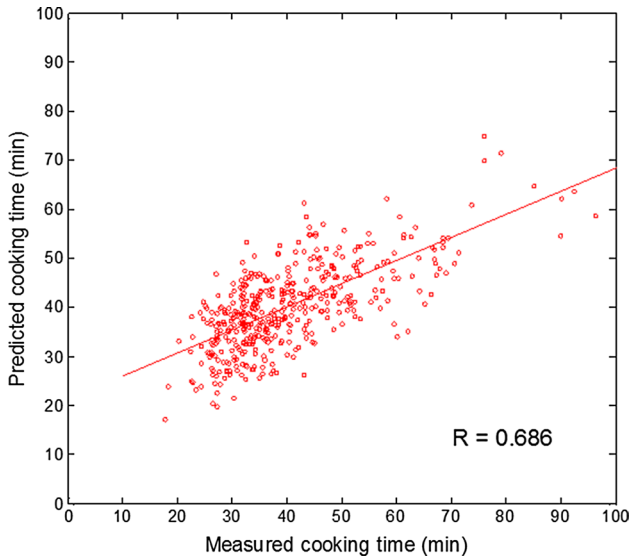

Fig. 3 Predictions for cooking time of dry beans using visible and near-infrared (Vis/NIR) spectroscopy. The best 37 selected features explained the variability of cooking time in beans with a correlation coefficient, $R = 0.686$

(ranging from 1012 to 2436 nm) as shown in Fig. 2. Figure 3 shows the predicted cooking time performance by the Vis/NIR data.

Fast cooking beans

The screening of the ADP for cooking time revealed five accessions that were especially fast cooking. Each of these accessions cooked in less than 27 min and the fastest of these cooked in less than 22 min (Table 6). In relation to the average cooking time of the entire set of 206 accessions, these fast cooking genotypes cooked at least 12 min

Table 6 Identification of the five fastest cooking accessions of the 206 Andean Diversity Panel accessions which cooked at least 12 min faster than the average cooking time in 2012 and 2013

ADP ID	Genotype	Origin	Source	Seed type	Cultivation status	Cook time (min) 2012	Min faster than avg[a]	Cook time (min) 2013	Min faster than avg.
ADP0367	G23086	Malawi	CIAT	Cranberry	Landrace	16.5	19.9	21.3	17.0
ADP0521	Cebo, Cela	Angola	Angola	Yellow	Landrace	20.8	15.6	26.7	11.6
ADP0469	PI527521	Burundi	US GRIN	White	Landrace	22.5	13.9	18.0	20.3
ADP0518	Mantega,blanca	Angola	Angola	Yellow	Landrace	23.9	12.5	22.7	15.6
ADP0452	Blanco Fanesquero	Ecuador	INIAP	White	Variety	24.5	11.9	24.1	14.2

[a] Determined by subtracting the avg. cooking time in a particular year from the cooking time of the accession of interest in the same year

faster in both field seasons (Table 6). Interestingly, 4 of the 5 accessions are landraces from various locations in Africa and one is a cultivar from South America. There are three different seed types represented in this group of five accessions: cranberry, yellow, and white (Table 6). The phylogenetic relationship of this set in relation to the entire set of 206 ADP accessions was evaluated with Nei genetic distance (Nei et al. 1983) from a set of nearly 5000 polymorphic SNPs. A neighbor joining tree developed with the data revealed that four of the five accessions cluster on the same branch of the tree, including ADP0367 from Malawi, ADP0518 and ADP0521 from Angola, and ADP0452 from Ecuador (Fig. 4). The genetic relatedness of these accessions suggests that they may share common genetic control for their fast cooking phenotype. ADP0469, a white seeded landrace from Burundi, is on a separate branch of the tree and, therefore, may have a unique genetic mechanism for fast cooking.

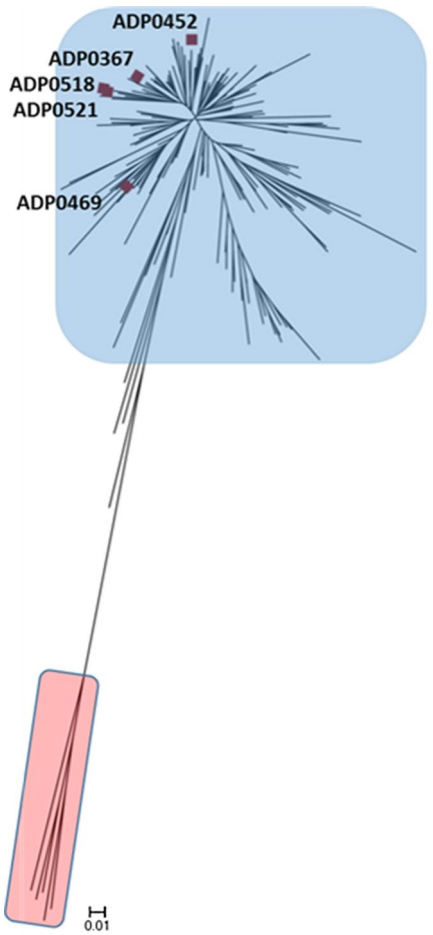

Fig. 4 Neighbor joining tree of 206 Andean Diversity Panel Accessions developed with 4935 SNP markers and using Nei Genetic Distance (1983). This tree is the consensus of 1000 bootstraps. A blue box surrounds the Andean germplasm and a *red box* surrounds the Middle American germplasm according to results by (Cichy et al. 2015). The location of the five fastest cooking ADP lines are highlighted with *red boxes* on the tree (color figure online)

Genome-wide association mapping

GWAS was conducted for water uptake during soaking and for cooking time. For both traits, there were significant SNP markers detected above the FDR cutoff (Table 7). For water uptake, significant marker trait associations were found on chromosomes Pv01, Pv03, Pv06, and Pv07 (Table 7; Fig. 5). The two associations on Pv01 were the most significant and also explained the most phenotypic variation with marker R^2 values of 14.6 and 25.1 %, respectively. The association on Pv03 explained 5.7 % of the phenotypic variation. The significant markers for water uptake detected on Pv06 explained from 4.6 to 8.2 % of the variation. The significant markers on Pv07 explained 13.3 % of the phenotypic variation for water uptake (Table 7). The significant SNP on Pv01, Pv03, and Pv06 have minor allele frequencies (MAF) of 0.04 or less. The two significant SNPs detected on Pv06 had much higher MAF (0.29–0.32) than those detected on Pv01, Pv03, and Pv06. Evaluation of the genomic region between these two SNPs on Pv06 revealed ten genes, including genes coding for a cystatin/monellin superfamily protein, tonoplast dicarboxylate transporter, and an acyl-CoA sterol acyl transferase 1.

Table 7 GWAS significant markers, genome position, *p* value, R^2, allele and allele frequency, and phenotypic effect associated with water uptake and cooking time

Trait	Marker	Chr	Position[a]	*p* value	R^2 (%)[b]	Allele	Effect	Obs
Water uptake (%)								
	ss715639380	1	36919960	1.83E−07	14.6	A	4.91	194
						G	0	6
	ss715640804	1	37661566	2.29E−11	25.1	A	24.80	195
						G	0	5
	ss715639608	3	33874326	7.89E−04	5.8	A	−8.36	192
						G	0	8
	ss715639606	3	33910221	7.89E−04	5.8	T	−8.36	192
						G	0	8
	ss715648493	6	25645359	9.89E−04	8.3	T	−1.68	142
						G	2.85	55
						K	0	2
	ss715645753	6	25750827	0.0027	4.6	A	−3.69	134
						C	0	63
	ss715644076	7	29497723	5.55E−07	13.3	A	3.72	194
						G	0	6
Cook time (min)								
	ss715649687	2	37679737	0.005	4.0	T	−16.9	197
						C	0	8
	ss715647434	2	38662923	0.00503	4.0	A	−17.6	199
						G	0	7
	ss715646000	2	48072125	0.00159	5.0	A	7.00	184
						C	0	22
	ss715646002	2	48100200	0.00159	5.0	T	7.00	184
						C	0	22
	ss715648837	3	832263	0.00345	4.3	T	−23.9	197
						C	0	9
	ss715650437	3	863021	0.00224	4.7	T	−21.6E	199
	ss715641024	6	7120266	3.99E−05	8.7	C	0	7
						A	14.51	198
	ss715640681	6	7351028	3.99E−05	8.7	G/Y	0	7/1
						T	14.51	198
	ss715640783	6	8538559	3.99E−05	8.7	C	0	7
						T	14.51	198
	ss715640782	6	8565591	3.99E−05	8.7	G	0	7
						T	14.51	198
	ss715642453	6	9204575	3.99E−05	8.7	C	0	7
						T	14.51	198
						C	0	7

[a] Position is based on the *P. vulgaris* reference genome (Schmutz et al. 2014)

[b] R^2 is the percent of phenotypic variation explained by the SNP marker

Significant SNPs associated with cooking time were found on chromosomes Pv02, Pv03, and Pv06 (Fig. 6). On Pv02, the significant SNPs mapped to two unique locations with different effects on cooking time. The first one (between ss715649687 and ss715647434) had an R^2 of 4 %. The major allele reduced cooking time by about 17 min and was found in 96 % of the diversity panel. For both of the significant SNP markers, the minor allele was found in five U.S. cranberry bean accessions with cooking times ranging from 44 to 83 min. There were a few candidate genes of interest that potentially are involved in determining cooking time. One candidate gene is Trichome Birefringence-like 13; in *Arabidopsis* this gene influences secondary cell wall structure and esterified pectin content

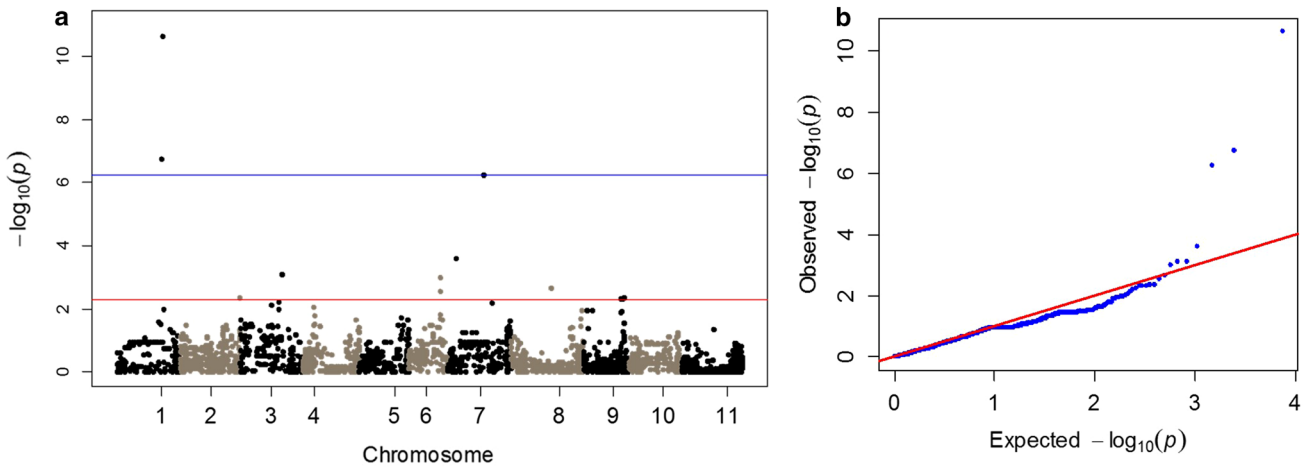

Fig. 5 **a** Manhattan plot of water uptake during 12 h. soak averaged across 2012 and 2013 of 206 Andean Diversity Panel accessions with 3741 SNP (>0.02 minor allele frequency) and mapping conducted with Multiple Linear Model using a kinship matrix and 2 principal components to account for population structure. The blue line is the False Discovery Rate (5.55×10^{-7}) threshold for significance and the red line is $p = 0.005$ threshold for significance. **b** QQ plot for the cooking time data with the model described above

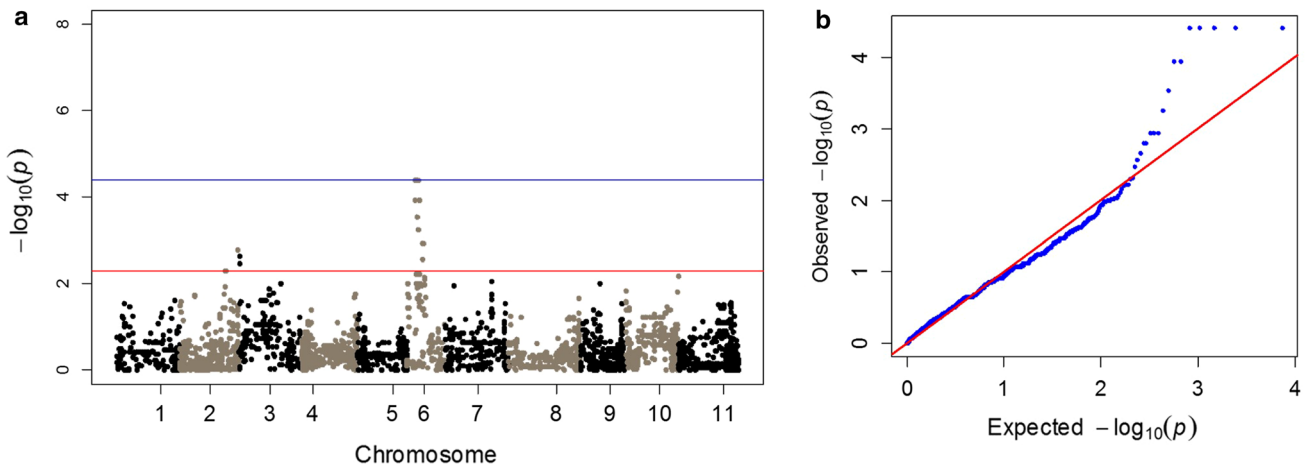

Fig. 6 **a** Manhattan plot of cooking time averaged across 2012 and 2013 of 206 Andean Diversity Panel accessions with 3741 SNP (>0.02 minor allele frequency) and mapping conducted with Multiple Linear Model using a kinship matrix and 2 principal components to account for population structure. The *blue line* is the False Discovery Rate (3.99×10^{-5}) threshold for significance and the red line is $p = 0.005$ threshold for significance. **b** QQ plot for the cooking time data with the model described above (color figure online)

(Bischoff et al. 2010). The second pair of SNPs (between ss715646000 and ss715646002) on chromosome Pv02 had an R^2 of 5 % (Table 7) and the major allele increased cooking time by 7 min (Table 7). The minor allele was found in 22 accessions of diverse types, including white, light red kidney, and yellow. The origin of these lines was also diverse, with some from Africa, Europe, North America, and South America. The two significant SNPs on chromosome Pv03 had R^2 values of less than 5 % (Table 7) and the major allele decreased cooking time by more than 20 min. The accessions with the minor allele were mostly of Caribbean origin. The most significant set of SNPs associated with cooking time were found on chromosome Pv06. This group of five SNPs was in linkage disequilibrium and explained 8.7 % of the phenotypic variation for this trait (Table 7) and the major allele which was found in 96 % of the accessions increased cooking time by 14.5 min. The minor allele for each of the SNPs in this block was detected in one of the five fastest cooking accessions in the panel (ADP0367), but was also detected in some lines with longer cooking times, including ADP0443 which cooks in 49 min.

Discussion

While substantial research has been conducted to understand the effect of the environment on cooking time in pulses, including dry bean, there has been limited assessment of genetic diversity and elucidation of genetic control of cooking time. In this screening of cooking time on 206 accessions, significant genotypic variability (5.5 fold) was observed. Dry beans come in many seed shapes, colors, and sizes with strong regional preferences for particular market classes (Voysest and Dessert 1991). The germplasm evaluated for cooking time included accessions from eight seed types of economic importance in Eastern and Southern Africa, Caribbean, North America, and South America. Within most of these market classes, a range of cooking times was observed and the fastest cooking lines cooked 20 min to more than 70 min faster than the slowest cooking lines. The diversity of cooking times within market classes is an important finding since breeding progress is typically more rapid when crossing within a market class.

Of particular interest are the white beans, which on average cooked faster that colored beans. The faster cooking time may be due to the lower levels of phenolic compounds in the seed coat as compared to colored seeds (Akond et al. 2011). Numerous studies have found a relationship between polyphenols and increased cooking time in beans (Srisuma et al. 1989; Stanley et al. 1990). The darker beans such as the red kidney and small red tended to take longer to cook on average than the lighter yellow and purple speckled. The cranberry bean, on the other hand, had one of the highest average cooking times and also the largest range in cooking time among accessions. This is interesting because the cranberry bean has a cream/white background with red mottling, so one would have expected it to be faster cooking if total polyphenolic compound levels are a genetic determinant of cooking time. Cranberry beans also have a tendency to darken within short periods of time in storage so their cream seed background turns brown. There is genetic variability in the extent of after darkening (Elsadr et al. 2011), which may have contributed to the wide contrast in cooking times among the different cranberry accessions. Studies are underway to profile the polyphenols in each of these 206 ADP accessions.

The significant association of cooking time with origin suggests that this trait is valued differently depending on the region where it is consumed. The two regions most represented were Africa with 93 accessions and North America with 68 lines. The 5-min difference in cooking time between the two regions suggests that in area where cooking is done with fuel wood, this trait may be more valued.

One area to follow up on is the role of gene pool in determining cooking time. This panel is largely Andean with only five Mesoamerican lines so it is difficult to make conclusions based on gene pool in this study. While the GWAS data presented for cooking time included the Mesoamerican lines, the analysis was also conducted with only the Andean beans of the ADP (data not shown) and patterns of significant SNPs associated with cooking time were similar with and without the inclusion of the Mesoamerican accessions.

Cooking time was weakly correlated with seed size, water uptake during soaking, and seed type. Previous studies have demonstrated relationships between cooking time and water uptake during soaking (Elia et al. 1997), while other studies found no correlation (Garcia et al. 2012). Elia et al. (1997) also found that higher tannin levels increased the cooking time. These previous studies support the correlation with cooking time and market class identified in the ADP.

The evaluation of Vis/NIR data and cooking time suggest this method could be used to quickly identify fast cooking bean lines based on characteristics of the dry seed. The number of selected features for the best predictive models is relatively large but not exceptional. Modeling the interaction of light with bean seeds for prediction mainly of internal quality parameters is a difficult task, which is affected by a number of factors. Consequently, a large number of variables in the best MLR models could be explained by the complexity of the bean microstructure (internal and in the first coat layers), physicochemical variations of the evaluated bean types, and/or seed curvature or unevenness on the measured surface of samples that affect the reflected light measured by the Vis/NIR spectrophotometer. The incorporation of the data for water absorption after soaking or water uptake ratio, seed type, and region of origin in the MLR model did not show improvements in the prediction of cooking time. Combinations of best Vis/NIR data and these parameters showed R values lower than 0.679.

Considering the complexity of the process involved in cooking beans, as well as the limitations of the reflectance mode to evaluate the core microstructure of the bean samples where drastic physicochemical and microstructural changes occur during cooking; overall results revealed promising opportunities for automatically measure cooking time using Vis/NIR techniques. Nonetheless, a better experimental setup for Vis/NIR analysis involving the transmittance mode or spectral imaging technique would be an interesting alternative for improving the accuracy of the prediction models.

The identification of fast cooking germplasm from diverse market classes has value for breeding. These accessions would also be excellent candidates to serve as parental lines in the development of recombinant inbred line populations for genetic studies to elucidate the mechanism of the fast cooking trait as a complement to GWAS. The fast cooking lines were under represented in the panel (only

2.4 % of the population) and if the phenotype is determined by rare alleles, they would not be easily detected with GWAS (Gibson 2012).

The GWAS results for water uptake are supported by previous research. QTL studies with *P. vulgaris* in an Andean x Mesoamerican cross of Xana x Cornell 49242 identified a QTL for water uptake (after 18 h. soak) on Pv03. Based on the location of the nearest marker on that map with known sequence (Bng21) the QTL was located to the same chromosome arm as the SNPs identified here for water uptake on Pv03 (Perez-Vega et al. 2010). The identification of a SNP for water uptake on Pv07 is of interest because the Asp gene is also located on chromosome Pv07. The Asp gene is responsible for the shiny seed coat observed in some *P. vulgaris* accessions. This trait has been shown to be associated with water uptake in black beans (Cichy et al. 2014). The location of the significant SNP on Pv07 (ss715644076) is roughly 27 Mb away from the estimated location of the Asp gene based on mapping studies (Cichy et al. 2014).

A few genetic studies have been conducted previously for cooking time. In an evaluation of a RIL population segregating for cooking time, the progeny of 3 generations (F6, F7, and F8) had three to one ratio of fast cooking to slow cooking, suggesting two genes controlling cooking time in this population (Jacinto-Hernandez et al. 2003). The research presented here found markers with small effects on cooking time (8.7 % and less), but these differences may partly be explained by the use of single parent crosses for QTL analysis vs association mapping. In another evaluation of cooking time on 140 F2:4 families segregating for cooking time, six QTL were identified for cooking time in a single environment. Three QTL were identified on chromosome Pv01 and three on Pv09 (Garcia et al. 2012), none of which appear to be similar to those presented here with GWAS.

Genes of interest on Pv06 near SNPs associated with cooking time include three UDP-glucosyl transferase genes. This gene family is involved in flavonoid biosynthesis and the development of pigments (Bowles et al. 2005) and are of interest because seed coat color (via market class) was correlated with cooking time and that the most significant NIR wavelength was 520 nm which is related to anthocyanins.

Other candidate genes for cooking time near the significant SNPs on Pv06 included two cation/H + exchanger genes, one homologous to AtCHX3 and one homologous to AtCHX4. The function of these genes in Arabidopsis has been related to the transport of Ca^{+2} (Manohar et al. 2011; Shigaki and Hirschi 2006). There is evidence that Ca^{2+} plays an important role in storage-induced increases in cooking time. A majority of calcium in seeds is stored with phytate and during storage phytate is broken down and free calcium is available to crosslink pectin in the cell wall, thereby strengthening cell wall and increasing the cooking time (Caffall and Mohnen 2009; Jones and Boulter 1983). In experiments growing peas (*Pisum sativum*) under different levels of Ca, K, and Mg, and P, it was revealed that peas cook faster the higher the ratio of monovalent to divalent cations in the fertilizer (Mattson et al. 1950).

In summary, this study identified wide variability for cooking time within dry beans of mostly Andean origin. Fast cooking accessions were identified and may be useful in breeding programs. Vis/NIR spectroscopy was shown to have potential as a method to predict cooking time on dry seed. GWAS identified regions on Pv01, Pv03, Pv06, and Pv07 related to water uptake during soaking and regions on Pv02, Pv03, and Pv06 related to cooking time. This research serves as an important foundation for further studies to understand the genetic control of cooking time in beans.

Author contribution statement KAC designed experiments, analyzed data, and wrote manuscript. JAW collected and analyzed data. FAM conducted Vis/NIR analysis with modeling and wrote manuscript.

Acknowledgments This work was supported in part by funding from the Norman Borlaug Commemorative Research Initiative (US Agency for International Development), by the US Department of Agriculture, Agricultural Research Service. The contents of this publication do not necessarily reflect the views or policies of the US Department of Agriculture, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. We also thank Yasmin Salat, Dennis Katuuramu, and Scott Shaw for assistance with cooking.

Conflict of interest The authors declare no conflict of interest.

References

Adkins E, Tyler E, Wang J, Siriri D, Modi V (2010) Field testing and survey evaluation of household biomass cookstoves in rural sub-Saharan Africa. Energy Sustain Dev 14:172–185

Akibode S, Maredia M (2011) Global and regional trends in production, trade and consumption of food legume crops. Department of Agricultural, Food and Resource Economics, Michigan State University

Akond A, Khandaker L, Berthold J, Gates L, Peters K, Delong H, Hossain K (2011) Anthocyanin, total polyphenols and antioxidant activity of common bean. Am J Food Technol 6:385–394

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (Methodological) 57:289–300

Bischoff V, Nita S, Neumetzler L, Schindelasch D, Urbain A, Eshed R, Persson S, Delmer D, Scheible W-R (2010) Trichome Birefringence and its homolog AT5G01360 encode Plant-specific DUF231 proteins required for cellulose biosynthesis in *Arabidopsis*. Plant Physiol 153:590–602

Bowles D, Isayenkova J, Lim E-K, Poppenberger B (2005) Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 8:254–263

Box GE, Cox DR (1964) An analysis of transformations. J R Stat Soc Series B (Methodological) 26:211–252

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Bressani R (1983) Research needs to up-grade the nutritional quality of common beans (*Phaseolus-vulgaris*). Qual Plant Plant Foods Human Nutr 32:101–110

Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr Res 344:1879–1900

Cichy K, Fernandez A, Kilian A, Kelly J, Galeano C, Shaw S, Brick M, Hodkinson D, Troxtell E (2014) QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris L.*). Mol Breed 33:139–154

Cichy KA, Porch T, Beaver JS, Cregan P, Fourie D, Glahn R, Grusak M, Kamfwa K, Katuuramu D, McClean P, Mndolwa E, Nchimbi-Msolla S, Pastor-Corrales MA, Miklas PN (2015) A *Phaseolus vulgaris* diversity panel for Andean bean improvement. Crop Sci. doi:10.2135/cropsci2014.09.0653

Coelho CMM, de Mattos Bellato C, Santos JCP, Ortega EMM, Tsai SM (2007) Effect of phytate and storage conditions on the development of the 'hard-to-cook' phenomenon in common beans. J Sci Food Agric 87:1237–1243

Drewnowski A, Rehm CD (2013) Vegetable cost metrics show that potatoes and beans provide most nutrients per penny. PLoS ONE 8:e63277

Elia FM, Hosfield GL, Kelly JD, Uebersax MA (1997) Genetic analysis and interrelationships between traits for cooking time, water absorption, and protein and tannin content of Andean dry beans. J Am Soc Hortic Sci 122:512–518

Elsadr HT, Wright LC, Pauls KP, Bett KE (2011) Characterization of seed coat post harvest darkening in common bean (*Phaseolus vulgaris L.*). Theor Appl Genet 123:1467–1472

Felix M, Gheewala SH (2011) A review of biomass energy dependency in Tanzania. Energy Procedia 9:338–343

Furst T, Connors M, Bisogni CA, Sobal J, Falk LW (1996) Food choice: a conceptual model of the process. Appetite 26:247–266

Garcia RAV, Rangel PN, Bassinello PZ, Brondani C, Melo LC, Sibov ST, Vianello-Brondani RP (2012) QTL mapping for the cooking time of common beans. Euphytica 186:779–792

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet 13:135–145

Grubbs FE (1950) Sample criteria for testing outlying observations. Ann Math Stat 21:27–58

Guthrie JF, Lin B-H, Frazao E (2002) Role of food prepared away from home in the American diet, 1977-78 versus 1994-96: changes and consequences. J Nutr Educ Behav 34:140–150

Hernández-Unzón HY, Ortega-Delgado ML (1989) Phytic acid in stored common bean seeds (*Phaseolus vulgaris L.*). Plant Foods Human Nutr 39:209–221

Jacinto-Hernandez C, Azpiroz-Rivero S, Acosta-Gallegos JA, Hernandez-Sanchez H, Bernal-Lugo I (2003) Genetic analysis and random amplified polymorphic DNA markers associated with cooking time in common bean. Crop Sci 43:329–332

Jones PMB, Boulter D (1983) The analysis of development of hardbean during storage of black beans (*Phaseolus vulgaris L*). Plant Foods Human Nutr 33:77–85

Liu K (1995) Cellular, biological, and physicochemical basis for the hard-to-cook defect in legume seeds. Crit Rev Food Sci Nutr 35:263–298

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Maes WH, Verbist B (2012) Increasing the sustainability of household cooking in developing countries: policy implications. Renew Sust Energ Rev 16:4204–4221

Manohar M, Shigaki T, Mei H, Park S, Marshall J, Aguilar J, Hirschi KD (2011) Characterization of *Arabidopsis* Ca2 +/H + exchanger CAX3. Biochemistry 50:6189–6195

Mattson S, Åkerberg E, Eriksson E, Koutler-Andersson E, Vahtras K (1950) Factors determining the composition and cookability of peas. Acta Agric Scand 1:40–61

Menendez A, Curt MD (2013) Energy and socioeconomic profile of a small rural community in the highlands of central Tanzania: a case study. Energy Sustain Dev 17:201–209

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170

Perez-Vega E, Paneda A, Rodriguez-Suarez C, Campa A, Giraldez R, Ferreira JJ (2010) Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 120:1367–1380

Rockland LB, Jones FT (1974) Scanning electron microscope studies on dry beans. Effects of cooking on the cellular structure of cotyledons in rehydrated large lima beans. J Food Sci 39:342–346

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MMS, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet 46:707–713

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52:591–611

Shigaki T, Hirschi KD (2006) Diverse functions and molecular properties emerging for CAX Cation/H + Exchangers in Plants. Plant Biol 8:419–429

Siddiq M, Uebersax MA (2012) Dry beans and pulses production and consumption-an overview. In: Siddiq M, Uebersax MA (eds) Dry beans and pulses production, processing and nutrition. Blackwell Publishing Ltd, Oxford, UK. doi:10.1002/9781118448298.ch1

Smith LP, Ng SW, Popkin BM (2013) Trends in US home food preparation and consumption: analysis of national nutrition surveys and time use studies from 1965–1966 to 2007–2008. Nutr J 12:45

Srisuma N, Hammerschmidt R, Uebersax M, Ruengsakulrach S, Bennink M, Hosfield G (1989) Storage induced changes of phenolic acids and the development of hard-to-cook in dry beans (*Phaseolus vulgaris var. Seafarer*). J Food Sci 54:311–314

Stanley D, Michaels T, Plhak L, Caldwell K (1990) Storage-induced hardening in 20 common bean cultivars. J Food Qual 13:233–247

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tuan YH, Phillips R (1992) Nutritional Quality of Hard-to-cook and Processed Cowpea. J Food Sci 57:1371–1374

Turner SD (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv (2014): 005165

Vindiola O, Seib P, Hoseney R (1986) Accelerated development of the hard-to-cook state in beans. Cereal foods world (USA)

Voysest O, Dessert M (1991) Bean cultivars: classes and commercial seed types. Common beans: research for crop improvement CAB International, Wallingford :119–162

Wang N, Daun JK (2005) Determination of cooking times of pulses using an automated Mattson cooker apparatus. J Sci Food Agric 85:1631–1635

Wang X, Hansen C, Allen K (2013) Identification of Anthocyanins Isolated from Black Bean Canning Wastewater by Macroporous Resin Using Optimized Conditions. Food Nutr Sci 4:174

Winham D, Webb D, Barr A (2008) Beans and good health. Nutr Today 43:201–209