Genome-wide Association Analysis for Drought Tolerance and Associated Traits in Common Bean

Valerio Hoyos-Villegas, Qijian Song, and James D. Kelly*

Abstract
A genome-wide association study explored the genetic basis of variation for drought tolerance and related traits in a Middle American diversity panel comprising 96 common bean (*Phaseolus vulgaris* L.) genotypes. The panel was grown under irrigated and rainfed conditions and single nucleotide polymorphism (SNP) data were used to explore the genetic diversity and ancestry of the panel. Varying levels of admixtures and distinctly divergent individuals were observed. Estimations of genome-wide heterozygosity revealed that, on average, greater diversity is present in individuals with Mesoamerica (3.8%) ancestry, followed by admixed individuals (2.3%). The race Durango had the lowest level of heterozygosity (1.4%). We report 27 significant marker–trait associations based on best linear unbiased predictors. These associations include seven markers for shoot biomass at harvest under irrigation and five markers under rainfed conditions on *P. vulgaris* (*Pv*) chromosome *Pv*11, two markers for shoot biomass at flowering under irrigation on *Pv*02 and *Pv*08, two markers for seed size under irrigated and rainfed conditions on *Pv*09, seven markers for lodging score under irradiation on *Pv*02 and *Pv*07, one marker for leaf elongation rate on *Pv*03 and one for wilting score on *Pv*11. Positional candidate genes, including *Phvu l.011G10270 0* on *Pv*11, associated with wilting, were identified. The SNP ss715639327 marker was located in the exon region of the *PvSIP1;3* gene, which codes for an aquaporin associated with water movement in beans. Significant quantitative trait loci identified in this study could be used in marker-assisted breeding to accelerate genetic improvement of drought tolerance in common bean.

Core Ideas
- GWAS was conducted to identify signals associated with drought tolerance in common bean.
- Significant signals for drought tolerance were identified.
- Population structure analysis of a panel of Mesoamerican genotypes revealed heterozygosity.

Common bean is one of the most important food legumes grown worldwide, with production figures exceeding 23 Tg. Most of the production (7 Tg) is concentrated in Latin America and eastern Africa (Broughton et al., 2003; Varshney et al., 2012), as the dry bean production figures reported by the Food and Agriculture Organization in Asia include largely *Vigna* rather than *Phaseolus* species (Akibode and Maredia, 2011). The highest per capita bean consumption is found in Rwanda and Burundi (40 kg person$^{-1}$ yr$^{-1}$), where common bean comprises 65% of total protein and 32% of energy consumed and provides important micronutrients such as Fe, Zn, thiamin and folic acid (Welch et al., 2000). In the United States, the common bean crop constitutes a commodity valued at approximately US$1 billion annually, with North Dakota and Michigan being the largest producers (Câmara et al., 2013).

Abbreviations: CMLM, compressed mixed linear model; DAP, days after planting; GAPIT, Genome Association and Prediction Integrated Tool; LD, linkage disequilibrium; MAF, minor allele frequency; *Pv*, *Phaseolus vulgaris* chromosome; QTL, quantitative trait locus; QTN, quantitative trait nucleotides; SNP, single nucleotide polymorphism.
Drought stress is a major limiting factor in common bean production. Drought events, both intermittent and terminal, affect over 60% of the annual dry bean production worldwide (White and Singh, 1991). In northeastern Brazil and the northern central highlands of Mexico, drought events occur annually and affect more than 1.5 million ha. The bean-growing regions of the western United States are also subject to drought conditions and thus irrigation is necessary to bring a crop to harvest, making this region one of the largest water consumers in the country (Muñoz-Perea et al., 2006). For this reason, a major goal of bean breeding programs has been to introduce drought tolerance into new cultivars to meet the ever-reducing water availability worldwide (Singh, 1995; Porch et al., 2009; Urrea et al., 2009; Builes et al., 2011).

Breeding for drought tolerance poses difficulties, primarily caused by large environmental and temporal variation, which make drought stress highly variable to observe (Rao, 2002). However, the development of drought-tolerant varieties is essential to enhance performance under stressed and non-stressed conditions (Miklas et al., 2006). The highest level of drought tolerance in dry beans has been reported in the race Durango, followed by the races Mesoamerica and Jalisco (Singh, 2007). Race Durango genotypes acquired drought tolerance due in part to their origins in the semiarid area of northern and central Mexico (Singh, 2007). Bean breeders have identified breeding lines from the race Durango that were superior to landraces from the same race as an outcome of double crosses between genotypes from Durango and Mesoamerica (Terán and Singh, 2002a). An example is the development of the drought-tolerant breeding line SEA 5 (Terán and Singh, 2002b), which has been widely deployed as a parent in many breeding programs (Mukeshimana et al., 2014a). Since seed yield is a priority in most breeding programs, previous work has suggested that the geometric mean between stress and nonstress treatments should be used as a selection criterion for performance. Geometric means and drought stress index can be used to select lines that have the highest yield performance under stress and nonstress conditions (Schneider et al., 1997). Detailed physiological studies have determined that variation for traits conferring drought tolerance is limited and indicate that the best method for selecting superior individuals under drought would be to find correlations between or among phenology, biomass production, and harvest index (Ramirez-Vallejo and Kelly, 1998). Other studies have suggested that surrogate measures of drought tolerance are leaf wilting and senescence; these traits were found to be correlated with leaf area index, biomass, and yield in common bean (Acosta Gallegos and Shibata, 1989). Recently, a study of a recombinant inbred line population derived from DOR364 × BAT477 tested in eight rainfall environments resulted in the detection of multiple quantitative trait loci (QTLs) in individual environments that were otherwise not detected when multiple QTL detection procedures were used (Asfaw et al., 2012). Other reports have suggested that the fraction of remobilized photosynthates from pods into seed is correlated with seed yield under stressed and nonstressed conditions. Photosynthetic remobilization is displayed in the Durango-derived breeding line SEA 5 and the Mesoamerica line G21212 (Asfaw et al., 2012; Araújo et al., 2014; Mukeshimana et al., 2014a).

Genome-wide association studies (GWAS) have rapidly become a powerful tool for the identification of candidate regions associated with quantitative traits. In common bean, this technology is showing promise in the identification of genomic regions associated with agronomic and phenological traits and symbiotic N₂ fixation (Kamfwa et al., 2015a, 2015b). With the recent release of the common bean genome (Schmutz et al., 2014), a set of genes linked with increased leaf and seed size was identified. The release has also provided a number of avenues for the study of the common bean genome and the implementation of new breeding technologies, such as genotyping arrays, as these provide a dense genome-wide set of SNP markers for more accurate exploration of the genomic regions responsible for a particular trait in common bean.

The objectives of this study were to explore the genetic variation present within a 96-entry diversity panel and use GWAS to identify candidate regions associated with drought tolerance traits and agronomic performance in common bean genotypes from the Middle American gene pool.

Materials and Methods

Germplasm

A North American diversity panel comprised of 498 individuals was assembled as part of the Bean Coordinated Agricultural Project from selections, breeding lines, and varieties made by the common (dry) bean breeding programs across the United States and Canada (http://www.beancap.org, accessed 10 Nov. 2016). The 96-genotype subset of Middle American genotypes used in the current study was selected on the basis of existing data and previous knowledge of the genotypes’ performance under rainfed or water-limited conditions, which identified them as drought-tolerant varieties, lines, or landraces. Additional information on the assembly of the 96-genotype panel can be found in Mukeshimana et al. (2014b).

Genetic Markers

The association panel was genotyped using the BARCBean6K_1 and BARCBean6K_2 BeadChips (Song et al., 2015; Schmutz et al., 2014). A dataset consisting of 96 accessions genotyped with 10,913 SNPs were obtained from the Soybean Genomics and Improvement USDA Laboratory (USDA–ARS, Beltsville Agricultural Research Center, Beltsville, MD). DNA was extracted as part of the Bean Coordinated Agricultural Project in facilities at North Dakota State University (Moghadam et al., 2016). A group of 3968 SNPs that did not have physical positions on the common bean genome were
eliminated. With the objective of obtaining more available markers, their flanking sequences were obtained using the dbSNP database (Sherry et al., 2001) and these sequences were aligned to the common bean genome using the Bowtie package (Langmead et al., 2009), which allowed for a total of 6865 SNP markers (Song et al., 2015). After this step, marker filtering involved eliminating markers with a minor allele frequency (MAF) of <5%, markers with >80% missing data, and identical markers. Identical markers were eliminated after calculating the percentage of identity among marker pairs and eliminating the marker with the greatest amount of missing data. At this point, a total of 3626 markers remained. Imputation of the remaining markers was done using BEAGLE (Browning and Browning, 2011). The missing data genotypes were equivalent to 4.2% before imputation, indicating an almost complete matrix at this stage.

Locations and Environment
Field trials were conducted at the Montcalm Research Center at Entrican, MI, in 2011 and 2013 and at the Saginaw Valley Research and Extension Center in Richville, MI, in 2012. Two-row field plots were planted at both locations and plots were 4.5 m in length with 0.5 m between rows. In 2012 and 2013, irrigated and rainfed plots with three replications each were planted but in 2011, only rainfed plots (with two replications) were planted because of seed quantity limitations. The irrigated and rainfed treatments were grown under an 8 by 12 lattice design.

To determine the extent of soil moisture depletion across the growing seasons, soil moisture sensors (IRROMETER Inc.) were installed at 20, 40, 70, and 100 cm of soil depth in two replications of each treatment in 2012 and 2013 and in rainfed conditions only in 2011. For simplicity, only the soil profile and replication averages within each year are shown (Fig. 1).

Phenotypic Data
Different measurements were performed over the growing season to study the yield, agronomic, and drought tolerance trait variation present in the association panel. The agronomic traits measured were days to flowering, days to maturity, shoot biomass at flowering and shoot biomass at harvest. Flowering date was measured as days after planting (DAP) when 50% of the plot showed flowers. Maturity date was determined as the number of DAP when the entirety of the plot had senesced and was considered to have reached harvest maturity. Lodging was scored on a scale from 1 = erect to 7 = lodged prostrate. Yield was determined by harvesting 4 m of both center rows, followed by open-air drying of the seed and seed moisture determination for correction to 18% moisture. Seed size was determined as 100-seed weight corrected to 18% moisture.

Wilting and Leaf Growth Rate
Greenhouse experiments were performed to assess the variation within the panel for the degree of wilting and the rate of leaf expansion under drought and well-watered conditions. To obtain an accurate assessment of the extent of wilting across the panel, two replications of every genotype were grown in small pots and allowed to grow under well-watered conditions for 3 wk. At this point, a dry-down period was initiated and wilting scores were measured daily for 20 d. Wilting was scored on a scale of 0 to 5, with 0 being no sign of wilting and 5 being completely wilted (Mukeshimana et al., 2014b). Leaf elongation was determined by tagging three young trifoliate leaves and measuring the length of the leaflets from the end of the petiole to the apex of the leaf. Measurements were taken daily until no further change in elongation was measured. Leaf elongation rate was calculated by solving for the slope of a linear equation where the dependent variable was time and the independent variable was leaf length.

Statistical Analyses
Linear Mixed Models
To minimize the effects of environmental and year-to-year variation, best linear unbiased predictions were used as phenotypic data input in the GWAS model and were calculated within treatments across years and across years and treatments. A mixed model in R using the package lme4 (Bates et al., 2014) was used to solve for the random effects (best linear unbiased predictions) of the model:

\[
Y_{ijk} = \mu + G_i + E_j + A_k + GE_{ij} + GA_{ik} + \varepsilon_{ijk}
\]

where \( m \) is the mean, \( G_i \) is the effect of the \( i \)th genotype (fixed), \( E_j \) is the effect of the \( j \)th treatment (fixed), \( A_k \) is the effect of the \( k \)th year (fixed), and \( \varepsilon_{ijk} \) is the error term (random). To address issues relating to small population size and detection power, we conducted simulations to determine the statistical power associated with different levels of Type I error in the simultaneous detection of 5, 15, and 30 quantitative trait nucleotides (QTN) with corresponding heritability estimates of 0.1, 0.5, and 0.9.

Inference of Population Structure
Determination of the number of subpopulations was performed using the console of the STRUCTURE 2.3.4 software package (Pritchard et al., 2000) using an admixture model for individual ancestry and allele frequency correlations. The markers used for this procedure were additionally filtered on the basis of pairwise linkage disequilibrium values greater than \( r^2 = 0.8 \) to comply with the requirement of the software that the markers used should be at or near Hardy–Weinberg equilibrium. The assumed number of subpopulations was simulated from \( k = 1 \) to \( k = 10 \) for an initial assessment of the most likely number of subpopulations with 10,000 iterations of burn-in followed by 50,000 Markov chain Monte Carlo iterations once the ideal number of subpopulations (\( K \)) was found by examining the optimal \( \Delta K \) value (Evanno et al., 2005) in STRUCTURE Harvester (Earl and vonHoldt, 2012). A
Fig. 1. Soil moisture and precipitation dynamics across the (a) 2011, (b) 2012, and (c) 2013 growing seasons at the Montcalm Research Farm in 2011 and 2013 and the Saginaw Valley Research and Extension Center in 2012. Daily values are the average across the soil profile from sensors placed at 20, 40, 70, and 100 cm in the soil profile, with two replications in each treatment.
total of $1 \times 10^5$ iterations of burn-in, followed by $5 \times 10^5$ Markov chain Monte Carlo iterations at the true value of $K$ was used to obtain the final population structure ($Q$) matrix and build a plot of mean likelihood values for each of the genotypes. The number of subpopulations and the replications within subpopulation were inputted into the Python script StrAuto version 0.3.1 (Chhatre, 2012), which was used to automate the simulation of different $K$ values in the STRUCTURE console.

Population structure was also studied by building a cladogram using a neighbor-joining clustering method in TASSEL version 5.0 (Bradbury et al., 2007) and by performing principal component analysis (Zhao et al., 2007) in the Genome Association and Prediction Integrated Tool (GAPIT, Lipka et al. (2012)). The number of principal components used was equal to the suggested number of subpopulations ($K$) described above by STRUCTURE. Relatedness was determined by calculating the kinship coefficient matrix ($K$), which was also estimated by GAPIT. The number of principal components used in GAPIT was determined by the number of subpopulations suggested by STRUCTURE. Heterozygosity ($H$) was also estimated to determine the level of gene diversity for every individual and to compare individuals across the diversity panel. Heterozygosity was estimated as:

$$H = 1 - \sum_{i=1}^{n} p_{ij}^2$$  \[2\]

where $p_{ij}$ is the sum of the frequency of the $i$th allele for the $j$th individual across all markers (Blair et al., 2009).

**Genome-wide Association Studies**

In GWAS, control of the population structure’s effects and unequal individual relatedness are two major causes of spurious associations. To solve this issue, we used the compressed mixed linear model (CMLM) method (Li et al., 2014) explained as

$$Y = X\beta + Z\mu + \epsilon$$  \[3\]

Based on the notation of Henderson (1984), $Y$ can be considered as the vector of observed phenotypes; $\beta$ is a vector of fixed effects, which includes the genetic markers, population structure ($Q$), kinship ($K$), and an intercept; $\mu$ is a vector of random additive genetic effects from multiple QTLs; $X$ and $Z$ are design matrices related to $\beta$ and $\mu$, respectively; and $\epsilon$ is the vector of residuals.

Upon calculation of the associations, the resulting $p$-values were corrected for multiple testing based on the false discovery rate (FDR) criterion as per Benjamini and Hochberg (1995). An false discovery rate-adjusted $q$ value of less than 0.05 was applied as the threshold for declaring an association as significant, which was done using the integrated functions within GAPIT. Once the corrected $p$ values were obtained, Manhattan and quantile-quantile plots were constructed. Linkage disequilibrium (LD) among or between marker pairs was estimated as the square of the correlation coefficient (Zhu et al., 2008) using the Synbreed package (Wimmer et al., 2012) in R, which can be used to invoke PLINK (Purcell et al., 2007) in the pairwiseLD function. The estimation of LD was used to examine haplotypes in the region after finding an association and to explore the intervening annotated genes from the reference genome for that region.

**Heritability**

Heritability was calculated across traits using mixed models. In summary, a genetic variance component was obtained from the genotype term of the model treated as random, one of the functions of GAPIT. Genetic variance in the case of a diversity panel is the portion of phenotypic variance attributable to familial relatedness estimated from all possible pairwise comparisons among genotypes (kinship). Heritability was estimated as

$$b^2 = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e}$$  \[4\]

where $\sigma^2_a$ is the additive genetic variance extracted from the variance component of the vectors $u$ and $e$ from

$$\text{Var} \begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$  \[5\]

where $G = \sigma^2_a$, $K$ is the kinship matrix, and $R = \sigma^2_e$, where $\sigma^2_e$ is the residual variance (Lipka et al., 2012).

**Results and Discussion**

**Environmental Conditions**

During 2011, 2012, and 2013, the 3 yr of field testing, 202, 241, and 184 mm of precipitation was recorded for the 4-mo period from planting to harvest (June–September). Precipitation was below the 30-yr average of 292 mm for the same months between 1951 and 1980 or when compared with the 335 mm from the 30-yr average between 1971 and 2000 according to the Michigan State University climatology records (http://climate.geo.msu.edu/climate_mि/stations/7227/, accessed 10 Nov. 2016). Each year showed distinct weather patterns and some differences were observed in terms of soil moisture dynamics. The year with the lowest precipitation and relatively high temperatures was 2013, as shown by the progressive decrease in the water tension readings from approximately 40 to 70 DAP, when a rain event occurred. The preceding years also showed some water deficit; however, precipitation was better distributed (Fig. 1). For example, in 2011, brief periods of dry-down were observed between 70 and 80 DAP and 90 to 100 DAP. However, these periods were relieved by rainfall before significant decreases in soil moisture could be observed. In 2011, a maximum temperature of
34°C was registered and a season average of 19.3°C was observed. In 2012, a maximum of 38°C and a season average of 18°C were registered. In 2013, a maximum of 35.6°C and a season average of 18°C were observed.

**Population Structure**

Population structure is an important covariate used in the CMLM method of association mapping to account for differentiation among panel groups and to avoid or minimize Type I errors. It is also important to study population structure in the context of genetic diversity and breeding to identify the genetic composition and relatedness of the individuals within the group. In our study, a clear subpopulation structure was observed among the individuals used in the panel. The subpopulation structure was confirmed by the suggested value of K in the population structure simulations and observation of neighbor-joining trees. The Evanno method suggested that the most likely number of subpopulations in the panel was K = 2, indicated by the largest change in ΔK among assumed K values (Evanno et al., 2005). Population structure can be observed in Fig. 2b. Based on the market classes of the genotypes and their clustering, the results suggest that the two subpopulations identified correspond to the races Mesoamerica and Durango, with admixed individuals between these two races (Singh et al., 1991). Approximately 27% of the genotypes were assigned to the race Mesoamerica, 34% were considered as admixture between races, and 40% were assigned to traditional Durango race genotypes grown in the western United States (Singh, 2007). Genotypes such as Mayflower, Midnight, Navigator, and Shania showed a probability of ancestry of 1 toward the Mesoamerica race. In contrast, Common Pinto, Gloria, NW-63, UI-114, UI-239, UI-537, Viva, and Yolano showed a probability of 1 toward the Durango race. A number of materials were found with probabilities of ancestry between 0.25 and 0.75 and were considered as admixed materials between the two races. Approximately equal admixture levels were observed for the genotypes GN9-4, F07-449-9-3, Matterhorn, Seafarer, Kodiak, Santa Fe, Buster, ND-307, Shiny Crow, La Paz, and NE2-09-3 (Fig. 2). The Type II growth habit (Singh, 1982) of many of these genotypes confirms the level of admixture that has occurred between the Durango and Mesoamerica races (Kelly, 2001), as opposed to the traditional Type III growth habit observed in the majority of the individuals observed to have more than 0.75 probability of ancestry belonging to the Durango race (Singh, 2007).

The race Mesoamerica is characterized primarily by the small-seeded black, small red, small white, Rosinha, and Carioca market classes (Beebe et al., 2000). We found that the group with an ancestry probability greater than 0.75 associated with race Mesoamerica contained all these market classes but did not contain any from the other races that are common with bean materials with the Type II growth habit (Diaz and Blair, 2006). In contrast, the race Durango is predominantly characterized by the Great Northern, Pinto, Bayo, Garrapato, and Ojo de Cabra market classes (Beebe et al., 2000). We found that the group denoted as primarily being associated with race Durango (more than 0.75 ancestry probability) contained mostly pinto and Great Northern genotypes, as these are the main seed classes grown in North America. In a similar study, Diaz and Blair (2006) observed an identical grouping of race structure within the Mesoamerican gene pool on the basis of microsatellite markers. They also found a high degree of gene flow between the Durango and Mesoamerica races in a number of the lines they evaluated. Blair et al. (2009) indicated that gene flow within the Mesoamerican gene pool is also significant, similar to findings in our study. They noted that individuals with a higher ancestry related to race Mesoamerica mainly comprise small-seeded genotypes with a Type II growth habit. In contrast, genotypes with ancestry in the Durango race are mainly represented by medium-seeded materials with a Type III growth habit, whereas the admixed materials exhibit a Type II growth habit. The admixed materials from North America result from breeding efforts to introduce the Type II habit from Mesoamerica into the traditional decumbent Type III Durango genotypes to improve plant architecture for disease avoidance and permit direct harvest (Kelly, 2001; Hoyos-Villegas et al., 2015).

The use of the two first principal components to account for population structure in the GWAS model covered 86% of the variance according to the cumulative eigenvalues reported by GAPIT. The identification of only two subpopulations in the panel results in a subtle level of population stratification for association studies (Zhu et al., 2008). This indicates that spurious marker–trait associations caused by population structure were minimized, despite the fact that 73% of the pairwise kinship values were below 0.6. From examination of the frequency distribution of kinship across the panel to determine the level of diversity under study, we found that only 27% of the kinship values were above 0.8, which suggests that the composition of the panel was diverse.

The self-pollinated nature of common bean results in high estimates of LD decay, assuming a threshold of \( r^2 = 0.1 \). Linkage disequilibrium decay can still reach up to 10 Mb in some chromosomes according to estimates using the panel used in this study (data not shown). This indicates that the identification of significant marker–trait associations with a relatively small sample size can still be found. Additionally, any significant signals found when using a diversity panel with a small sample size are indicative of alleles being present at a high frequency when detected under a MAF larger than 5%, provided that the effects of multiple testing, population structure, and kinship are minimized.

Different patterns in heterozygosity levels were observed across races. We found a 0.5% difference between Mesoamerica and the admixed materials and a 0.9% difference between the admixed group and the Durango materials. However, there is a higher number
Fig. 2. Results of the structure analysis performed on the diversity panel used in this study: (a) neighbor-joining dendrogram of the diversity panel; (b) results from population structure simulations at $K = 2$. Genotypes are on the $x$-axis and the probability of ancestry from the races Mesoamerica and Durango are on the left and right $y$-axes, respectively. Lines indicating different levels of admixture among populations are drawn at 0.25, 0.5, and 0.75. Genotypes with probabilities of ancestry from Mesoamerica (>75%), admixtures (75–50 and 50–25%), and Durango (>75%) are colored in blue, black, green, and red, respectively.
of individuals in the admixed group with large heterozygosity percentages than in the two other groups, but the heterozygosity mean is reduced by the admixed group, which contains more individuals than the races Mesomaerica and Durango (Supplemental Fig. S1). Single locus heterozygosity in the panel averaged 2.3%, which is consistent with the values reported by Kwak and Gepts (2009) and is expected, given the self-pollinated nature of the species. The estimations of genome-wide heterozygosity (Supplemental Fig. S1) revealed that, on average, greater diversity is present in individuals with more than 75% Mesoamerica (3.8%) ancestry, followed by the individuals considered to be admixed (2.3%). Individuals with more than 75% Durango ancestry had the lowest level of heterozygosity (1.4%). When individual heterozygosity was studied, the five individuals that presented the highest level within the Durango race ranged from USRM-20 (7.1%) to UI-537 (2.0%). Correspondingly, the five individuals with the lowest heterozygosity values had a value of 0.6%. Within the materials grouped as admixed between Mesoamerica and Durango, the five individuals with the highest level of heterozygosity ranged from GN9-1 (15.4%) to ND-307 (3.8%). Within this same group, the five individuals with the lowest heterozygosity had values from 0.6 to 0.7%. The final group, comprised of individual groups within Mesoamerica, indicated that the five individuals with the highest level of heterozygosity were 115M (Black Rhino) (35.8%), and other genotypes from A-55 (10.7%) to IBC301-204 (6.9%). Likewise, the five individuals with the lowest heterozygosity had a value of 0.8% (Supplemental Fig. S1). It is interesting to observe a large difference in heterozygosity in the breeding line 115M (Black Rhino), which showed a 25% difference in heterozygosity compared with GN9-1, the individual with next highest value (10.7%). In a study to identify QTLs associated with canning quality in the black bean marker class, Wright and Kelly (2011) developed a recombinant inbred line population between the cultivar Jaguar and the breeding line 115M. The 115M line was developed at CIAT (Cal, Colombia) from an inbred backcross between Tacana (Lopez-Salinas et al., 1997) and G24423, a Colombian selection from wild germplasm identified from a unique genetic marker pattern in prior diversity studies (Tohme et al., 1996). The genetic background of the wild breeding line may provide an explanation for the distinct level of heterozygosity calculated from genome-wide SNPs when compared with the other cultivated materials present in the diversity panel. 115M was derived from fewer generations of selfing than some of the historical landraces such as common pinto or the more commonly grown cultivars, which would have had greater opportunity to eliminate heterozygosity from their genomes through repeated selfing during decades of seed multiplication.

**Marker–Trait Associations**

A total of 27 significant marker–trait associations were found among the traits evaluated and the final set of markers used in this study. These associations are shown in Fig. 3. Significant markers were found for shoot biomass at harvest under rainfed and irrigated conditions, shoot biomass at flowering under irrigated conditions, 100-seed weight under rainfed and irrigated conditions, and lodging score under irrigated conditions. Significant associations were found for the greenhouse variables, including wilting score at 28 DAP and leaf elongation rate. No significant markers were detected for yield in this study. The minimum, means, and maximum values of the phenotypic data used for these analyses are shown in Tables 1 and 2. Specifically, seven SNPs were found on chromosome Pv11 that were significant for shoot biomass at harvest under irrigated conditions and five additional SNPs for shoot biomass were also detected at harvest under rainfed conditions on Pv11. Shoot biomass at flowering presented two significant SNPs on Pvo2 and Pvo8. The 100-seed weights under irrigated and rainfed conditions showed two significant associated SNPs, both on Pvo9. Lodging score showed seven significant SNPs on Pvo2 and Pvo7. Leaf elongation rate in the greenhouse showed one significant SNP on Pvo3 in addition to one significant SNP on Pvo11 for wilting score at 28 DAP.

To alleviate concerns that the small panel size would result in false associations, we conducted simulations on the SNP dataset using CMLM analysis to determine the resulting statistical power to detect a range of QTNs with varying heritability values. When the Type I error rate increased the probability of detection increased as well, almost at a 1:1 ratio when the trait was controlled by 30 QTNs and had a heritability of 0.1 (Supplemental Fig. S2). A trait with a heritability of 0.9 that is controlled by 15 or 30 QTN has a lower detection power than a trait controlled by 5 QTNs and a heritability of 0.5. In real terms, a trait controlled by 5 QTN and a heritability of 0.5 would indicate that it would be detected around 50% of the time when the rate of Type I errors is below 0.1. Another output of the simulations was the ability to detect changes in the R² of the CMLM GWAS model when the peak SNPs were removed. Across all traits, R² values were reduced by more than half their original value when SNPs with significant effects were removed, suggesting that only valid associations were being detected in our study.

**Shoot Biomass at Harvest**

The false discovery rate adjusted p-values determined a significant (p < 0.05) signal for shoot biomass (Table 1) at the time of harvest under both irrigated and rainfed conditions (Fig. 3a and 3b). Under both conditions, a significant association was found to peak at the SNP ss715648770 (Fig. 3a and 3b), which is located at 8.1 Mb on Pv11. This SNP had a MAF of 0.4 in the population. Heritability was 33.8% under rainfed conditions and 56.9 under irrigated conditions. Linkage disequilibrium values showed that two SNPs were in strong LD with ss715648770 and the SNP markers ss715640673.
Fig. 3. Manhattan plots and respective quantile-quantile (Q-Q) plots of the phenotypic traits evaluated to have significant associations: (a) shoot biomass at harvest under rainfed conditions, (b) shoot biomass at harvest under irrigated conditions, (c) shoot biomass at flowering under irrigated conditions, and (d) lodging under irrigated conditions.
and ss715649670 had $r^2$ values of 0.91 and 0.94, respectively. Within this 0.4-Mb region, 31 genes were found (the list is shown in Supplemental Table S1). However, the flanking sequence of the SNP ss715648770 aligns to the intronic region of the locus Phvul.011G085100, annotated as “charged multivesicular body protein.” The signal for shoot biomass at harvest under irrigated conditions detected by the signal on the SNP ss715648770 had an $R^2$ value of 0.2 with an effect of $-0.23$ kg m$^{-2}$, whereas the $R^2$ value for shoot biomass at harvest under rainfed conditions was 0.18. These $R^2$ values indicate that the amount of variation explained by this association ranged from 18 to 20% depending on the soil moisture availability. Seven SNPs were found to be significant under irrigated conditions (ss715648770, ss715646312, ss715640673, ss715646313, ss715646311, ss715649670, and ss715650173), whereas five SNPs were significant under rainfed conditions. Among these two groups, the markers overlapped under irrigated conditions, with the exception of the SNPs ss715640673, ss715646313, ss715649670, and ss715646310. Under irrigated conditions, the sum of the effect of all the significant SNPs was $-0.63$ kg m$^{-2}$ and 0.8 kg m$^{-2}$ for all the SNPs with a significant positive effect. Under rainfed conditions, the sum of the SNPs with a significant positive effect was 2 kg m$^{-2}$, whereas the only SNP with a significant negative effect was ss715648770 with $-0.52$ kg m$^{-2}$.

<table>
<thead>
<tr>
<th>Race</th>
<th>Trait</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Rainfed</td>
<td>Irrigated</td>
<td>Rainfed</td>
</tr>
<tr>
<td>Mesoamerica</td>
<td>Shoot biomass at harvest, kg m$^{-2}$</td>
<td>Min. 1.05</td>
<td>0.58</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 1.71</td>
<td>1.11</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 4.41</td>
<td>1.71</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>Shoot biomass at flowering, kg m$^{-2}$</td>
<td>Min. 5.31</td>
<td>29.67</td>
<td>33.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 12.11</td>
<td>49.89</td>
<td>55.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 23.67</td>
<td>77.13</td>
<td>77.35</td>
</tr>
<tr>
<td></td>
<td>100-seed weight, g</td>
<td>Min. 17.57</td>
<td>18.33</td>
<td>18.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 24.14</td>
<td>21.62</td>
<td>22.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 34.26</td>
<td>26.78</td>
<td>27.74</td>
</tr>
<tr>
<td></td>
<td>Lodging score (1–7)†</td>
<td>Min. 1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 2.20</td>
<td>2.48</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 4.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Admixture</td>
<td>Shoot biomass at harvest, kg m$^{-2}$</td>
<td>Min. 0.81</td>
<td>0.71</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 1.55</td>
<td>1.09</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 4.54</td>
<td>1.55</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>Shoot biomass at flowering, kg m$^{-2}$</td>
<td>Min. 4.00</td>
<td>29.85</td>
<td>37.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 14.66</td>
<td>56.77</td>
<td>60.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 28.18</td>
<td>94.80</td>
<td>101.75</td>
</tr>
<tr>
<td></td>
<td>100-seed weight, g</td>
<td>Min. 20.53</td>
<td>16.82</td>
<td>19.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 40.20</td>
<td>34.52</td>
<td>36.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 50.24</td>
<td>44.56</td>
<td>45.03</td>
</tr>
<tr>
<td></td>
<td>Lodging score (1–7)†</td>
<td>Min. 1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 2.77</td>
<td>3.46</td>
<td>3.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Durango</td>
<td>Shoot biomass at harvest, kg m$^{-2}$</td>
<td>Min. 0.81</td>
<td>0.52</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 1.69</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 5.43</td>
<td>2.28</td>
<td>2.41</td>
</tr>
<tr>
<td></td>
<td>Shoot biomass at flowering, kg m$^{-2}$</td>
<td>Min. 3.20</td>
<td>27.85</td>
<td>34.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 16.19</td>
<td>56.54</td>
<td>59.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 29.82</td>
<td>86.40</td>
<td>88.25</td>
</tr>
<tr>
<td></td>
<td>100-seed weight, g</td>
<td>Min. 29.25</td>
<td>28.20</td>
<td>30.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 39.75</td>
<td>34.97</td>
<td>37.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 50.15</td>
<td>43.37</td>
<td>46.39</td>
</tr>
<tr>
<td></td>
<td>Lodging score (1–7)†</td>
<td>Min. 2.00</td>
<td>1.00</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean 4.20</td>
<td>4.80</td>
<td>4.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Max. 5.00</td>
<td>5.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>

|        |        |        |              |              |              |              |              |
|        |        |        |              |              |              |              |              |
|        |        |        |              |              |              |              |              |

† Lodging score: 1 = erect; 7 = lodged prostrate.
Shoot Biomass at Flowering
Two significant associations were found for shoot biomass at the time of flowering under irrigated conditions. The signals were found on the SNPs ss715646929 and ss715639951 with MAFs of 0.23 and 0.06, respectively (Fig. 3c). Heritability was 54.9%. These markers are located at 46,685 and 56.1 Mb on Pv02 and Pv08, respectively. The $R^2$ for these two markers was around 0.20 for each of the associations and their effects were 0.41 and 0.16 kg m$^{-2}$ for the SNPs ss715646929 and ss715639951, respectively. No other markers were found in strong LD ($>0.5$) with these two SNPs. However, SNP ss715639951 is located on the gene Phvul.008G247200, which is annotated as “NB-ARC domain-containing disease resistance protein.” The 0.06-Mb region between the SNP in highest LD with ss715639951 contains six genes (Supplemental Table S2).

Aboveground biomass is related to crop yield in that some portion of biomass is the source of photosynthesize for the developing seed. In terms of stress tolerance, aboveground biomass at the time of harvest is an indicator of the ability of the plant to maintain growth and development over stress periods of varying duration. Not all the aboveground biomass is translocated into the harvestable product; the efficiency and extent with which this occurs is directed by flexible source–sink relationships. To date, the genetic architecture of these processes has not been completely understood (Wallace et al., 1993a, 1993b).

Crop biomass is well known to be closely related to crop transpiration and this balance is very sensitive to water status. In the short term, a reduction in leaf area will minimize water loss by reducing the water potential gradient between the leaves and the soil, which is similar to stomatal closure. Mechanisms to reduce water use for later stages of plant development are needed over longer periods of time (Tardieu, 2006). However, water conservation comes at a cost for the plant because of reduced C gain and photosynthesis rates. This cost is explained by the well-documented relationship between light interception and biomass production (Monteith and Moss, 1977), a phenomenon also documented in common bean (Beebe, 2012). A clear balance therefore exists between the reduction of severe stress conditions and water conservation for continued growth and development. When considering all the strategies conferring drought tolerance in crop plants, many of these will ultimately involve the regulation of water loss through the leaves while maintaining an acceptable level of biomass accumulation that translates into yield (Blum, 2011).

Previous studies have reported the importance of biomass production under stress and nonstress conditions in common bean. For example, Ramirez-Vallejo and Kelly (1998) found a strong correlation between shoot biomass and yield under stressed and nonstressed conditions in a group of genotypes contrasting for response under drought stress. Asfaw et al. (2012) showed a moderate phenotypic correlation between shoot biomass and grain yield under both drought stress and nonstress conditions and identified the QTL Cbm3.1 for shoot biomass on Pv03. The heritability estimate for shoot biomass reported by (Asfaw et al., 2012) was 0.36, which is in agreement with heritability estimates of 0.56 and 0.33 for irrigated and rainfed conditions, respectively, found in our study.

Carbohydrate allocation in common bean undergoes a major shift at the onset of flowering. At this point, the expanding population of small pods becomes a sink for the rest of the biomass. This is important in that the amount of biomass accumulated during the vegetative stages will establish the yield potential, as dry matter accumulation rates from sources such as leaves and stems will be significantly reduced (Hay and Porter, 2006) on flowering. Under drought stress, vegetative growth rates are affected, which implies that under vegetative-stage drought stress, the extent of achievable yield will be limited even if soil moisture is available for reproductive structure development to occur normally.

Lodging Score
Seven SNPs were significantly associated with lodging (Table 1) under irrigated conditions. These included four markers on Pv02, with ss715647096 being the most significant marker, which is located at 43.4 Mb on Pv02, and three markers on Pv07, with ss715646517 located at 46.1 Mb being the most significant SNP on Pv07 (Fig. 3d). These markers ranged in MAF from 0.13 to 0.25 and their $R^2$ values were all around 0.07. The markers ss715647096 and ss715647094 on Pv02 had an effect of 0.75 and ss715647095 and ss715647092 had an effect of −0.75. The markers ss715646517 and ss71564518 on Pv07 had an effect of −0.6, whereas the marker ss715646516 had an effect of 0.6. A total of four markers were in LD higher than 0.5 with the SNP ss715647096, which resulted in an 11.2-Mb region with 962 genes (Supplemental Tables S3a and S3b) and several significant molecular function gene ontology-enriched terms. On Pv07, the SNP ss715646517 had three other markers with a LD higher than 0.5 and this resulted in a 0.1-Mb region with nine genes (Supplemental Table S4). The sequence involving ss715647096 on Pv02 is located in the gene Phvul.002G269200, annotated as “histone H3.” The sequence involving ss715646517 on Pv07 is located on an exonic region of the gene Phvul.007G222200, annotated as “serine/threonine protein phosphatase.”

Previous studies have reported QTLs for lodging in Pv02 and Pv07. In a GWAS study of a larger Middle American panel, Moghaddam et al. (2016) identified a candidate gene at 45.6 Mb on Pv07, which was homologous to the SARI gene that controls stem thickness. In a different study to detect QTLs associated with white mold tolerance, Mkwalla et al. (2011) screened two inbred backcross line populations derived from the black bean Tacana as the recurrent parent and found a QTL for lodging in separate populations in linkage group B7 (Pv07) indicated by the simple sequence repeat markers F1R1.275 and FSR5.225, with $R^2$ values of 0.48 and 0.28, respectively. In a pinto × navy bean cross, Miklas et al. (2007) found a QTL for lodging in linkage group B2 (Pv02),
indicated by the marker AFLP10, with an additive effect of −0.3 lodging score units. Kolkman and Kelly (2003) reported the markers acgctt239 and 107.1200 to be linked with QTLs for lodging in Linkage Groups two and seven (Pv02 and Pv07), respectively, with mean effects of −0.3 and −0.4 lodging score units. In combined analyses, the amount of variance explained by the QTL in their study (0.07 and 0.09) is similar to that reported in this study. It is important to note that although all of these studies identified QTLs associated with a lodging phenotype, the primary result of Kolkman and Kelly (2003) was the identification of QTLs associated with avoidance to white mold disease. To ensure white mold disease development for screening purposes, the evaluation was conducted under supplemental irrigation conditions that favored the development of the infection. It is therefore interesting to note that our GWAS signals for lodging were only found under irrigated conditions, indicating the importance of this condition for the expression of the phenotype. In common bean breeding, lodging and irrigation are often connected when dealing with the fungal pathogen white mold [Sclerotinia sclerotiorum (Lib.) de Bary]. Breeding for cultivars that cope with this disease is often done through the selection of individuals with canopy architectural traits that prevent the formation of a microclimate that is favorable for disease development, particularly under irrigated conditions or in high rainfall years (Hoyos-Villegas et al., 2015). Heavy lodging increases disease pressure because of increases in humidity and reductions in temperature within the plant canopy.

Changing the plant architecture was one of the objectives for implementing ideotype breeding in common bean. The introduction of a modified ideotype (Adams, 1982) as a Type II growth habit cultivar over the original Type I determinate habit (Adams, 1973) of common bean has had a significant impact on seed yield and has changed how breeders visualize a successful cultivar (Kelly, 2001). The yield advantage of modifying the plant architecture was mainly attributed to a longer seed filling period (Izquierdo and Hosfield, 1983). The lack of symmetry among plant parts, combined with yield component compensation (Adams, 1967), pleiotropy, and genetic background, are factors that limit breeding progress for yield. However, the decision to breed for more biomass partitioning into the stem for improved stem strength to prevent lodging can be dependent on the target environment and therefore has not always been identified as a priority. For instance, the Type II growth habit in the semiarid conditions of the western United States has often been recognized as a disadvantage in terms of yield performance when compared with rainfed environments in the Midwest (Kelly et al., 1998).

Hundred-Seed Weight

The same two SNPs were significantly associated with 100-seed weight (Table 1) under both irrigated and rainfed conditions (Fig. 4a and 4b). Heritability was 93.4% under irrigated conditions and 89.9% under rainfed conditions. The markers ss715646851 and ss715646847 were located at 16,599 and 16.6 Mb, respectively, on Pv09, with the former being at the peak of the GWAS signal. These SNPs were estimated to have a MAF of 0.39 and the model $R^2$ value for the association was 0.04 for each SNP or 4% of the genetic variation, with effects of −3.0 and 3.0 g per 100 seeds under rainfed conditions and −2.6 and 2.6 g per 100 seeds under irrigated conditions for the SNPs ss715646851 and ss715646847, respectively. When pairwise LD estimations were made for ss715646851, 23 markers were found to have an LD of 0.5 or higher. The two markers in highest LD with ss715646851 were ss715646852 (0.94) and ss715646847 (0.93), which cover a 0.6-Mb region with 46 genes (Supplemental Table S5a and S5b). Two molecular function gene ontology terms (transferase activity, transferring phosphorus-containing groups and phosphotransferase activity, alcohol group as acceptor) were found as enriched in this region. Previous reports have also identified QTL for seed size on Pv09. Blair et al. (2006) identified the QTL SW9.1 using an advanced backcross analysis of a BC$_2$F$_3$ population derived from a cross of a cultivated Andean line with a wild bean accession. The same SW9.1 QTL with the nearest marker, BMd21, was found to be linked with QTLs for total seed yield and number of pods per plant. A BLAST search of this marker revealed that its sequence is located on the gene Phvul.009G116700 and is approximately 1 Mb downstream from SNP ss715646851. The LD analysis showed that the SNP ss715646447 is in high LD ($r^2 = 0.7$) with ss715646851 and is 0.1 Mb from Phvul.009G116700. Mkwaila et al. (2011) also identified a QTL for seed weight on Pv09 with an $R^2$ of 0.49 and similar additivity of 0.5 g per 100 seeds in an inbred backcross population between a cultivated Middle American and a wild bean accession. The seed size of a particular market class is important for consumer acceptance, so it is partly dependent on the maintenance of a consistent range in this trait. Market classes in the Mesoamerican genepool are often recognized as being small (17–22 g per 100 seeds) to medium (32–40 g per 100 seeds) but seed size variation is common under abiotic stress conditions such as drought or heat stress. Commercial classes are often referred to as being dependent on the development of an acceptable cultivar that is adapted to a particular environment and has traits that are desirable by growers, processors, and consumers (Siddiq and Uebersax, 2013). However, breeders are often limited by the amount of progress that can be made in other desirable traits because of the narrow quality limits that a cultivar must maintain to be accepted into the marketplace. For example, early generation testing and selection based on yield components have been adopted to improve seed yield in common bean; however, these have often met with limited success because of the large variation among market classes in seed size and pod number, which restricts improvement in yield caused by compensation among yield components (Kelly et al., 1998; Kelly and Cichy, 2013). Limitations associated with growth habit and seed size are also common and
Fig. 4. Manhattan plots and respective quantile–quantile (Q-Q) plots of the phenotypic traits evaluated to have significant associations: (a) 100-seed weight under irrigated conditions, (b) 100-seed weight under rainfed conditions, (c) leaf elongation rate, and (d) wilting score 25 days after planting.
are a cause for a lack of acceptance by both breeders and growers (Brothers and Kelly, 1993). Having markers that would allow the efficient selection of materials within a particular seed size range before flowering and maturity would be beneficial for a breeding program. Selection before flowering would save time in the identification of elite parents for crossing and selection before maturity would allow the screening of a larger number of early generation materials to be selected before advancing to field trials. Given the effect of the marker–trait association reported above, a potential exists for establishing such a tool. More associated loci and marker validation would be necessary to implement this. Further, the integration of the markers ss71564073 and ss715646847 as part of a selection system for seed size could be integrated into a processing quality index to facilitate selection of quality traits that will be essential in future bean cultivars (Hosfield, 1991; Kelly and Cichy, 2013).

Leaf Elongation Rate

The ss715640038 SNP associated with the leaf elongation rate trait (Table 2) was located at 2.8 Mb on Pv03 and had a MAF of 0.05 (Fig. 4c). The $R^2$ for this SNP was 0.43 and it had an effect of 0.12 cm d$^{-1}$. One other SNP was in high LD with this marker. This was ss715640477, with an $r^2$ of 0.79, which led to delimiting a 0.4-Mb region with 34 genes (Supplemental Table S6). Cell growth and division depend largely on turgor and cell wall extensibility and are founded on a classical model known as the Lockhart equation (Lockhart, 1965). The cell expansion rate is equal to $m(P - Y)$, where $m$ is cell wall extensibility, $P$ is turgor pressure, and $Y$ is the value below $P$ at which no cell wall growth occurs. Drought-limited cell wall expansion implicates a reduction in turgor and the loss of cell wall extensibility (Granier and Tardieu, 1999) before leaf photosynthesis is inhibited. Costa França et al. (2000) studied four bean cultivars from Brazil under water-limited conditions to assess growth parameters such as leaf relative growth rate as adaptive mechanisms. They reported different rates of leaf dehydration rates in some cultivars when compared with known drought-sensitive cultivars. Significant differences in growth rate among common bean genotypes have also been reported in their response to drought stress (Gebezhuhu, 2006; Lizana et al., 2006; Rosales et al., 2012). A significant association with leaf elongation rate suggests that there is potential for the exploitation of this trait for the improvement of drought tolerance in common bean. Since variation exists within the cultivated common bean germplasm, its genetic structure can be characterized and utilized in breeding efforts, along with other tolerance mechanisms.

Wilting Score

The GWAS procedure for wilting score (Table 2) at 28 DAP resulted in one significant marker association at the SNP ss715639678, which is located at 46.8 Mb on Pv11 (Fig. 4d). The $R^2$ for this association was 0.29 with an effect of −0.5 score units, indicating that this marker explains 29% of the variation and causes a one-half reduction in wilting score on a 0 to 5 scale. Heritability for wilting score was 26.3%. Two other SNPs, ss715639326 and ss715640758, were found to be in high LD with ss715639678, with $r^2$ values of 0.61 and 0.77, respectively. This led to defining a 38.6-Mb region with 1131 genes. In addition, gene ontology enrichment analysis revealed 19 biological processes and 30 molecular functions that were significantly associated (Supplemental Tables S7a–S7c). The sequence involving ss715639678 aligns with the exonic region of the gene Phvul.011G191800 of the common bean genome and is annotated as “leucine-rich repeat-containing protein.”

There is ample knowledge of differential responses to water limiting conditions and wilting in other grain legumes (soybean [Glycine max (L.) Merr.] and cowpea [Vigna unguiculata (L.) Walp.]) (Fletcher et al., 2007; Sadok and Sinclair, 2009, 2010; Abdel-Haleem et al., 2012). Simple evaluations can be used as the basis for identifying the underlying plant adaptation to drought to identify germplasm for use in introgression efforts. The US soybean breeding community has been successful in its efforts to identify slow-wilting genotypes. The basis for the development of drought-tolerant genotypes are the plant introductions PI 416937 and PI 471938 which have slower wilting under water limited conditions than other lines. Slow wilting is a water conservation trait that permits transpiration rates to continue increasing despite atmospheric vapor pressure deficits of ~2.1 kPa or

<table>
<thead>
<tr>
<th>Race</th>
<th>Trait</th>
<th>Value</th>
<th>Greenhouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesoamerica</td>
<td>Leaf elongation</td>
<td>Mean 0.13</td>
<td>Mean 0.13</td>
</tr>
<tr>
<td></td>
<td>rate (cm d$^{-1}$)</td>
<td></td>
<td>Max. 0.58</td>
</tr>
<tr>
<td>Wilting score at 25 DAP</td>
<td>Min. 0.00</td>
<td></td>
<td>Min. 1.00</td>
</tr>
<tr>
<td>(0–5)†</td>
<td></td>
<td></td>
<td>Max. 2.00</td>
</tr>
<tr>
<td>Admixture</td>
<td>Leaf elongation</td>
<td>Mean 0.13</td>
<td>Mean 0.13</td>
</tr>
<tr>
<td></td>
<td>rate (cm d$^{-1}$)</td>
<td></td>
<td>Max. 0.58</td>
</tr>
<tr>
<td>Wilting score at 25 DAP</td>
<td>Min. 1.00</td>
<td></td>
<td>Min. 1.00</td>
</tr>
<tr>
<td>(0–5)†</td>
<td></td>
<td></td>
<td>Max. 2.00</td>
</tr>
<tr>
<td>Durango</td>
<td>Leaf elongation</td>
<td>Mean 0.00</td>
<td>Mean 0.00</td>
</tr>
<tr>
<td></td>
<td>rate (cm d$^{-1}$)</td>
<td></td>
<td>Max. 0.00</td>
</tr>
<tr>
<td>Wilting score at 25 DAP</td>
<td>Min. 0.00</td>
<td></td>
<td>Min. 1.00</td>
</tr>
<tr>
<td>(0–5)†</td>
<td></td>
<td></td>
<td>Max. 2.00</td>
</tr>
</tbody>
</table>

† Wilting was scored on a scale of 0 to 5, with 0 being no sign of wilting and 5 being completely wilted (Mukashimana et al., 2014b).
‡ DAP, d after planting.
higher (Sadok et al., 2012). Since limited work has been conducted to survey wilting across the common bean germplasm, a comparative mapping study between the two species would provide the opportunity to exploit synteny (McClean et al., 2010) among the two genomes and to further exploit the present sizeable disparity between genomic resources in soybean and common bean.

A recent bioinformatic analysis of the common bean genome by Ariani and Gepts (2015) revealed 41 aquaporin-encoding (Gomes et al., 2009) genes from five subfamilies (PIPs, TIPs, NIPs, SIPs, and XIPs). Interestingly, examination of the genes in their study and the genes found in the LD interval reported in Supplemental Tables S7a to S7c revealed one common locus (Phvul.011G102700) on Pv11. This gene (PvSIP1;3) is annotated as a small and basic intrinsic protein 1A family aquaporin. The gene PvSIP1;3 was reported to be a plasma membrane-localized aquaporin in a study using the WoLF PSORT algorithm (Horton et al., 2007). It is important to note that the SNP ss715639327 is located in the exon region of PvSIP1;3 and within the LD range of the peak marker (ss715639678) associated with the wilting score reported in our study.

Conclusions

The objective was to conduct a GWAS of a panel of diverse Middle American bean genotypes previously selected on the basis of their drought tolerance to identify marker–trait associations. A number of associations were found based on best linear unbiased predictions from 3 yr of field studies, after analyzing a series of shoot- and root-related traits. We identified signals in shoot biomass at harvest, total shoot biomass at flowering, 100-seed weight, lodging score, leaf elongation rate, and wilting at 25 DAP. The amount of genetic variance explained by many of these signals ranged from 4 to 29%, indicating that many of these associations have significant effects and the genomic regions that are associated with these traits require further investigation. The population structure analyses revealed that the diversity panel was comprised of two subpopulations, based on their ancestry. Analysis of the genes found in the LD interval reported in our study and the genes in their study revealed one common locus (Phvul.011G102700) on Pv11. This gene (PvSIP1;3) is annotated as a small and basic intrinsic protein 1A family aquaporin. The gene PvSIP1;3 was reported to be a plasma membrane-localized aquaporin in a study using the WoLF PSORT algorithm (Horton et al., 2007). It is important to note that the SNP ss715639327 is located in the exon region of PvSIP1;3 and within the LD range of the peak marker (ss715639678) associated with the wilting score reported in our study.

Supplemental Material

Supplemental materials provide additional information on the molecular function and biological processes of candidate genes associated with biomass, flowering, lodging, seed weight, leaf elongation, and wilting traits.

Acknowledgments

We acknowledge support from the National Institute of Food and Agriculture (USDA) under Award no. 2010-85117-20570, the Agriculture and Food Research Initiative Project no. 2009-01929, and from Michigan State University AgBioResearch and the USDA–ARS. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the USDA.

References


