A *Phaseolus vulgaris* Diversity Panel for Andean Bean Improvement

Karen A. Cichy,* Timothy G. Porch, James S. Beaver, Perry Cregan, Deidre Fourie, Raymond P. Glahn, Michael A. Grusak, Kelvin Kamfwa, Dennis N. Katuuramu, Phil McClean, Eninka Mndolwa, Susan Nchimbi-Msolla, Marcial A. Pastor-Corrales, and Phillip N. Miklas

**ABSTRACT**

Dry beans (*Phaseolus vulgaris* L.) of the Andean gene pool, including red mottled, kidney, cranberry, and yellow seed types are important in Africa and the Americas. Andean dry bean breeding gains have lagged behind those of Mesoamerican beans. This difference may result from a narrower genetic base in the Andean gene pool and reduced breeding efforts. The objective of this research was to establish, genotype, and phenotype a panel of bean germplasm to be used for Andean dry bean breeding. An Andean diversity panel (ADp) was assembled, consisting of 396 accessions and including important cultivars, breeding lines, and landraces that originate mostly from Africa, the Caribbean, and North and South America. The panel was genotyped using the Illumina BARCBean6K_3 SNP BeadChip. The population contained two subgroups: Andean and Mesoamerican bean germplasm. The ADp was comprised of 349 Andean, 21 Mesoamerican, and 26 Andean–Mesoamerican admixed accessions. Most admixed lines came from Africa (12 accessions) and the Caribbean (five accessions). Association mapping was conducted for determinacy. Significant single-nucleotide polymorphism (SNP) trait associations were found on chromosome Pv01, with the most significant SNP marker being 3.1 kb from the *Terminal Flower 1 PvTFL1y* gene. The ADp was evaluated for numerous traits in field trials in the United States and Africa. Variability was found for resistance to rust, angular leaf spot and common bacterial blight diseases; tolerance to low soil fertility; cooking time; and other traits that can be used to improve Andean bean germplasm for Africa and the Americas.

**Abbreviations:** ADP, Andean diversity panel; ALS, angular leaf spot; CBB, common bacterial blight; GWAS, genome-wide association study; PCA, principal components analysis; RCBD, randomized complete block design; SNP, single-nucleotide polymorphisms.
is created by a bottleneck that reduced diversity in wild Andean beans long before domestication (Bitocchi et al., 2013; Schmutz et al., 2014). Andean beans were probably taken to Europe and were taken to Africa from there, possibly by Portuguese traders or by other European colonial countries beginning in the 16th century (Gepts and Bliss, 1988). Most of the Andean beans in the United States were probably brought by European immigrants (Gepts and Bliss, 1988; Gepts et al., 1988).

Dry beans from both gene pools have commercial importance today. Andean beans are preferred in some regions for their larger seed size. Andean beans are medium (25–40 g per 100 seeds) or large (>40 g per 100 seeds), whereas Mesoamerican beans are small (<25 g per 100 seeds) or medium-sized. Andean beans have been grouped into three races: Peru, Chile, and Nueva Granada, with the latter being the most widely grown around the world (Singh et al., 1991a). Although production of Mesoamerican beans predominates in agriculture systems in North and Central America, Andean bean production predominates in parts of Africa, Europe, and South America. In Europe, 75% of the bean germplasm is of Andean origin (Logozzo et al., 2007). In the African continent, half of the beans produced are from the Andean gene pool. However, in East and Southern Africa, Andean beans are preferred and an estimated 73 to 83% of beans are Andean in origin (Bellucci et al., 2014; Gepts et al., 1988).

In the United States, red kidney beans are the most important Andean type but from 2010 to 2012, they were just 6% of overall production. In spite of their minor status, they are a high value crop and they fetch prices as much as 40% higher than the commonly produced Mesoamerican bean seed types (Wells et al., 2014).

Genetic improvement of Andean beans has lagged behind improvement of Mesoamerican beans. This is true in Latin America, where from 1930 to 1993, only 25% of germplasm released through hybridization were Andean in origin (Voysest et al., 1994). In the United States, there has also been less effort in Andean bean breeding and there have only been modest breeding gains for yield, which has resulted in a reduction of acreage in Andean classes (Kelly and Cichy, 2012). The difference between the actual yield increase of 1.5% per year on farm and the 2.1% potential yield increase per year was most probably caused by more intensive crop management rather than breeding, as Andean beans in the United States are more heavily managed than Mesoamerican beans (Vandemark et al., 2014). The major breeding successes in Andean beans in North America include improvements in earliness and canning quality and the transfer of halo blight tolerance or resistance into kidney beans (Copeland and Erdmann, 1977; Kelly et al., 1998).

In contrast to Andean bean production in the United States, dry beans in Africa are typically grown on small farms with limited fertilization and other inputs. Soils in common bean production areas in Africa are also often depleted in N, P, or both (Graham et al., 2003). Soil N is limited on 50% and 60% of bean production land in Eastern and Southern Africa, respectively. Soil P is deficient on 65% and 85% of the land where beans are produced in Eastern and Southern Africa, respectively (Kimani et al., 2001). In Sub-Saharan Africa, dry bean yields average 0.62 tons ha⁻¹ according to 2006–2008 data and the 0.3% per year rate of yield increase is much lower than that of the United States (Akibode and Maredia, 2011). Some of the major Andean bean breeding successes in Africa include resistance to angular leaf spot (ALS) (caused by Pseudocercospora griseola), rust resistance, increased seed micronutrient concentrations, and tolerance to low soil fertility (Aggarwal et al., 2004; Beebe 2012; Blair et al., 2010; Liebenberg et al., 2005, Liebenberg and Pretorius, 2010).

There is a unique opportunity to create synergy between the breeding efforts in the Americas and in Africa for Andean beans to broaden the genetic base of Andean beans and to increase yield potential, improve disease resistance, and develop tolerance to low fertility for bean growing regions in Africa and the Americas. Previous diversity studies have characterized Andean bean germplasm both genotypically and phenotypically (Beebe et al., 2001). Many studies have focused on genetic diversity within a particular region, such as the Americas (Singh et al., 1991b), Africa (Asfaw et al., 2009; Blair et al., 2010), Europe (Angioi et al., 2010; Logozzo et al., 2007), or North America (McCLean et al., 1993). Diversity analyses also have the potential to be useful for breeding, especially as a way to identify promising parental lines. The goal of this research was to assemble a set of dry bean landraces and improved varieties for global Andean bean improvement, referred to as the ADP. The genotypic and phenotypic diversity of the ADP is described, as well as its usefulness for association mapping.

**MATERIALS AND METHODS**

**Plant Material**

A group of 396 *P. vulgaris* accessions was assembled into a panel to facilitate genetic improvement of beans of the Andean gene pool. These accessions are collectively referred to as the ADP. The ADP includes 120 lines from the United States National Plant Germplasm Collection, representing accessions collected in East Africa, 87 lines from U.S. bean breeding programs, 74 lines from CIAT in Cali, Colombia assembled from African breeding programs, 43 accessions from the CIAT Germplasm Collection, 28 lines and landrace cultivars used as checks or parents from African breeding programs, 15 lines from an Angolan *Phaseolus* collection, 11 lines from a Caribbean *Phaseolus* collection, 10 lines from Canadian breeding programs, and eight lines from Ecuadorian breeding programs.
Genotypic Evaluation

DNA was extracted from newly emerged trifoliate leaves using a hexadecyltrimethyl ammonium bromide method (Rogers and Bendich, 1985). The DNA concentrations were estimated by measuring the optical density at A260 and A280 wavelengths with a Nanodrop Spectrophotometer (ND–8000; NanoDrop Products, Wilmington, DE). The ADP was genotyped with the Illumina (Illumina Inc., San Diego, CA) BARCBean6K_3 SNP array (Song et al., 2015), which permits the analysis of 5398 SNP markers distributed across the 11 pairs of common bean chromosomes. The BACBean6K_3 BeadChips were scanned with the Illumina BeadStation 500G. Single-nucleotide polymorphism calling was conducted with the genotyping module V2011.1 of GenomeStudio software (Illumina Inc.).

Data Analysis

The program STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to determine the number of subpopulations in the ADP. The program uses Bayesian clustering based on molecular marker data to determine the population structure (Pritchard et al., 2000). One of the assumptions of the analysis is that markers are ‘loosely’ in linkage equilibrium (Pritchard et al., 2000). A subset of 392 SNP markers in linkage equilibrium was identified. Linkage disequilibrium was tested using GGT version 2.0 (van Berloo, 2008). Linkage disequilibrium was determined by the equation $D = D^2/p(1-p) q(1-q)$, where $p$ and $q$ are allele frequencies at two loci and $D$ is the deviation (Hill and Robertson, 1968), with $D$ greater than 0.8 considered in linkage disequilibrium.

The entire set of 5398 SNP markers was filtered to remove monomorphic SNPs and any SNPs with a minor allele frequency of 5% or less. The remaining set of 3385 SNP markers was used to evaluate the population structure via principal component analysis (PCA) using a correlation matrix in the program TASSEL version 4.0. The same set of SNPs was used to develop a neighbor-joining tree via the parsimony substitution model in TASSEL version 4.0. The tree graphic was developed using FIG TREE version 1.4.0 (Rambaut, 2012). The graphical display of population structure by region of origin was developed using the DISTSTRUCT software package (Rosenberg, 2004).

Diversity estimates, including major allele frequency, gene diversity, heterozygosity, and polymorphism information content were determined using PowerMarker version 3.25 (Liu and Muse, 2005). Major allele frequency is a measure of the proportion of lines with the most common allele for each SNP marker (Weir and Cockerham, 1996). The gene diversity is the probability that any two alleles chosen at random from within the population are different and is defined in Liu and Muse (2005).

Heterozygosity is the proportion of heterozygous individuals for each SNP (Weir and Cockerham, 1996). Polymorphism information content is a measure of polymorphism and is used to assess the usefulness of a molecular marker for linkage analysis (Botstein et al., 1980).

For the diversity analyses, 4935 SNP markers were used, which is the number that remained after filtering the entire set to remove markers with more than 10% missing data and those that were monomorphic in all 396 ADP lines. Each analysis was conducted on individual SNP markers and average diversity estimates were determined for germplasm subsets of interest.

A genome-wide association study (GWAS) for plant determinacy was conducted using TASSEL version 4.0 (Bradbury et al., 2007) on 374 ADP lines. The 4935 SNP markers described above were filtered to remove any SNPs with a minor allele frequency <0.05. After filtering, 3385 SNPs remained for GWAS analysis with a mixed linear model. Population structure was accounted for in the model with PCA and five principal components were included, which explained 51.3% of the variance (Table 1). A kinship matrix ($\Phi$) was also included in the association analysis to account for relatedness. The following mixed linear model equation was used: $y = X_0 + P\beta + \Phi_1 u + e$, where $y$ is phenotype, $X$ is SNP, $P$ is the PCA matrix, $\Phi$ and $e$ represent fixed effects, $\Phi$ is the relative kinship matrix (value) and $e$ is for residual effects. The cutoff used for significant SNP markers for determinacy was the Bonferroni correction $P = 1.48 \times 10^{-5}$ (for $\alpha = 0.05$ and 3385 SNPs).

Phenotypic Evaluation

Determinacy

Plant determinacy was determined from 374 ADP lines grown at the Agricultural Research Council Experimental Station in Cedara, South Africa. The trial was planted in January 2014 with three replications in a randomized complete block design (RCBD). Plants were classified as either determinate, with the main stem ending in a terminal flower bud, or indeterminate, where the flower bud was not terminal.

Common Bacterial Blight

A set of 215 ADP lines was planted at the Zambia Agricultural Demonstration Center at the University of Zambia in Lusaka, Zambia. The trial was planted in December 2013 with two replications in a RCBD. The panel entries ADP0665 (USWK-CBB–17) and ADP0676 (CEL.R.K) were used as the resistant and susceptible checks, respectively (Miklas et al., 2006). Common bacterial blight (CBB) disease, caused by Xanthomonas axonopodis pv. phaseoli, was prevalent at the site where the ADP was planted. The experimental site at the University of Zambia is in the region of Zambia that tends to have high CBB pressure.

<table>
<thead>
<tr>
<th>PC</th>
<th>Eigenvectors</th>
<th>Individual contribution</th>
<th>Cumulative contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1241.9</td>
<td>36.7</td>
<td>36.7</td>
</tr>
<tr>
<td>2</td>
<td>213.4</td>
<td>6.3</td>
<td>43.0</td>
</tr>
<tr>
<td>3</td>
<td>99.5</td>
<td>2.9</td>
<td>45.9</td>
</tr>
<tr>
<td>4</td>
<td>94.4</td>
<td>2.8</td>
<td>48.7</td>
</tr>
<tr>
<td>5</td>
<td>88.3</td>
<td>2.6</td>
<td>51.3</td>
</tr>
</tbody>
</table>

Table 1. Principal components and contributions for 396 accessions of the Andean diversity panel of common bean using 3385 SNP markers.
because of higher temperature and rainfall. The 2013 season was not an exception. The disease pressure was uniform across the experimental field, as reflected by the consistent scores obtained for resistant and susceptible checks in both replications. These natural infections were used to evaluate the ADP entries for their reaction to the CBB pathogen at the flowering and pod-filling growth stages. The reaction of the plant canopy to CBB for each genotype was evaluated based on a 1 to 9 scale (van Schoonhoven and Pastor-Corrales, 1987), where 1 = immune and 9 = very susceptible. The scores were grouped into three categories: 1–3 for resistant, 4–6 for intermediate, and 7–9 for susceptible.

**Angular Leaf Spot and Rust Diseases**
A set of 410 ADP lines were grown in Cedula, South Africa. The trial was planted in January 2014 in a RCBD with three replications and evaluated under field conditions for reaction to rust (caused by *Uromyces appendiculatus*) and ALS disease. A set of 12 check cultivars, six Andean and six Mesoamerican that were known for their reaction to the local isolates of the ALS and rust pathogens were planted in each of the replications of the ADP. In Cedara, rust and ALS disease occurred naturally and were widespread throughout the field.

For both diseases, the severity (area of the plant affected by the disease causing organisms) was evaluated in March 2014 using a 1 to 9 scale (van Schoonhoven and Pastor-Corrales, 1987) as described above. Each disease assessment was conducted on mature plants with pods at different stages of development.

**Root Rot**
A set of 246 ADP lines was grown in Isabela, Puerto Rico, in a field that has been in common bean production since 1971. The Mesoamerican small red bean germplasm, ‘TARS-LFR1’, was included in the experiment as a root rot-resistant check (Porch et al., 2014a). The trial was planted in December 2012 in an RCBD with two replications. The most prevalent soil-borne fungal pathogens in this nursery are *Fusarium solani* f. sp. *phaseoli*, *Macrophomina phaseolina*, and *Sclerotium rolfsii*, causing southern blight. Other fungi, including *Pythium aphanidermatum* and *Pythium graminicola* were identified with less frequency (Porch et al., 2014b). Low fertility conditions, specifically low N, were also prevalent, since fertilizer is not applied to this nursery. Seed yield was measured at harvest.

**Low Fertility**
A set of 270 ADP lines were tested for performance in low-fertility soil (<8 mg kg⁻¹ P and <40 kg ha⁻¹ N) on the Sokoine University Agricultural Farm in Morogoro, Tanzania. These accessions included ADP0097 (BILFA 4), a low-fertility-tolerant check (Hillocks et al., 2006). The trial was planted in April 2013 in a RCBD with two replications. Yield (kg ha⁻¹) was estimated from the harvest of individual one-row plots, 8 m in length and spaced 0.75 m apart.

**Drought**
A set of 144 ADP lines was planted at the Washington State University Research Farm in Othello, WA, during the 2012 growing season using a split-plot design with two replications and with treatment as the main plot and ADP accessions as the subplots. Drought stress was imposed by stopping furrow irrigation applications at flowering (50% of plants with at least one open blossom). There was no precipitation in Othello, WA, after flowering so the drought-stressed plants received no water, whereas the nonstressed plants received sufficient irrigation to achieve optimal yields. Plots consisted of four rows of each accession, with 0.56-m spacing between rows and a row length of 3 m. The two center rows were harvested for yield. Seed yield was determined as the geometric mean in kg ha⁻¹. The drought intensity index during the trial was 0.40 and indicates moderate drought. The drought intensity index is defined as: 1 – X_I (average trial seed yield under drought)/X_M (average trial seed yield under normal water; nonstressed).

**Seed Iron Concentration**
A set of 143 ADP lines was grown in Prosser, WA, in 2012, as described above. Dried seed samples (approximately 50 g) were ground to a fine powder using a Cyclone Sample Mill (Udy Corp.; Fort Collins, CO) with a stainless steel grinding ring. A minimum of two subsamples (~0.5 g dry weight) of each ground sample were digested and processed for elemental analysis as previously described (Farnham et al., 2011).

**Cooking Time**
A set of 240 ADP lines grown in 2012 and 2013 in a RCBD with two replications at the Montcalm Research Farm in Entrican, MI, was evaluated for cooking time. Cooking time was determined on each entry within 6 mo of harvest from seed stored in ambient conditions as follows: a sample of 25 seeds was soaked for 12 h in distilled water, weighed to determine water uptake, and cooked with a Matson bean cooker (Customized Machining and Hydraulics Co., Winnipeg, Canada) in boiling water. The optimum cooking time was recorded as the time it took for 80% of the plungers to pierce the seeds (Wang and Daun, 2005).

**Iron Bioavailability**
A subset of 204 of the cooked ADP lines described above for cooking time was evaluated for iron bioavailability. An *in vitro* caco-2 cell line assay was used to determine ferritin formation in the harvested caco-2 cell monolayers and ferritin formation was used to quantify Fe bioavailability (Glahn et al., 1998).

**Database**
A custom database was established to facilitate access to the phenotypic and genotypic data from the multiple projects and research groups using the ADP. The database was established using a LAMP (Linux, Apache, MySQL, PHP) software platform that is maintained by the USDA-ARS in Mayaguez, PR, and is housed on a server at the University of Puerto Rico, Mayaguez. Data include photos of seeds and plant rows in the field of individual ADP lines, basic agronomic and marker information on individual ADP lines, SNP haplotypes based on data collected from the BARCBean6K_3 SNP Illumina BeadChip, and data from field trials conducted on the ADP in the United States, Africa, Latin America, and the Caribbean. Andean diversity panel lines originating from the CIAT or USDA Genetic Resource Collections are hyperlinked to the
CIAT and Germplasm Resources Information Network databases, respectively. Data can be accessed through exporting data from whole trials or specific traits across trials to an Excel (Microsoft, Redmond, WA) spreadsheet format. The database will be publicly accessible at http://arsftbean.uprm.edu/bean/ (accessed 15 May 2015).

RESULTS
Genetic Diversity of the ADP
The ADP is a group of 396 P. vulgaris accessions with nearly equal numbers of landraces (201 accessions) and improved lines (195 accessions). The collection primarily consists of accessions of African (222 accessions) and North American (95 accessions) origin (Table 2). The Andean market classes of major importance in Africa and North America are represented in the panel, including kidney, red mottled, cranberry or sugar, and yellow types. There are also non-commercial seed types including brown and gray (Table 2).

Based on the STRUCTURE results, shown as ΔK (Fig. 1), the ADP is comprised of two sub populations, which correspond to the two gene pools: Andean and Mesoamerican. Most of the lines (349 accessions) clustered within the Andean gene pool. Twenty-one accessions clustered within the Mesoamerican gene pool and 26 accessions were an admixture of 10 to 90% of the two gene pools. The Mesoamerican lines included 16 landraces and five improved lines. Most of the Mesoamerican lines came from Africa (11 accessions) and the Caribbean (seven accessions) (Fig. 2) and the most represented seed types were red mottled and yellow. The admixed lines included 10 landraces and 16 improved lines. Most of these lines also came from Africa (12 accessions) and the Caribbean (five accessions) and the most represented seed types were red mottled, yellow, and black. Within the Andean lines, there was no strong clustering into subpopulations, but a further division into K = 3 was conducted into Andean Group 1 and Andean Group 2 (Fig. 2, Supplemental Table S1). A tendency was observed for those lines with the highest identity in Andean Group 1 to be kidney types and for those of Andean Group 2 to be red mottled types (Supplemental Table S1). A PCA was also conducted to evaluate population structure in the ADP. The first principal component accounted for 36.7% of the variance and separated the Mesoamerican and the Andean germplasm into the same clusters as were found with STRUCTURE (Fig. 3). The first principal component also separated the admixed lines from the Mesoamerican and Andean groups. The second principal component only accounted for 6.3% of the variance and separated the Andean germplasm such that a small subset of 14 accessions clustered together (Fig. 3, upper right corner). These are all improved determinate kidney lines from North America. The genotypes in this cluster all have ‘Great Northern 1’ (Mesoamerican) in their pedigree. Most have it by way of ‘Montcalm’, a cross between Great Northern 1 (Mesoamerican) and a red kidney bean. Great Northern 1 was used as a parental source for halo blight resistance (Copeland and Erdmann, 1977).

A neighbor-joining tree of the Andean ADP genotypes was produced to characterize the relatedness of the germplasm from different origins (Fig. 4, Supplementary Table 1). The African and North American accessions

Table 2. Classification of 396 Andean diversity panel common bean accessions by growth habit, origin, cultivation status, and seed type.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Number of lines</th>
<th>Seed type</th>
<th>Number of lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth habit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determinate</td>
<td>230</td>
<td>Beige</td>
<td>10</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>166</td>
<td>Black</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brown</td>
<td>11</td>
</tr>
<tr>
<td>Origin</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>222</td>
<td>Cranberry</td>
<td>56</td>
</tr>
<tr>
<td>Asia</td>
<td>6</td>
<td>Dark red kidney</td>
<td>24</td>
</tr>
<tr>
<td>Caribbean</td>
<td>26</td>
<td>Gray</td>
<td>4</td>
</tr>
<tr>
<td>Central America</td>
<td>6</td>
<td>Jacob’s cattle</td>
<td>6</td>
</tr>
<tr>
<td>Europe</td>
<td>9</td>
<td>Light red kidney</td>
<td>43</td>
</tr>
<tr>
<td>North America</td>
<td>95</td>
<td>Pink mottled</td>
<td>5</td>
</tr>
<tr>
<td>South America</td>
<td>32</td>
<td>Purple mottled</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purple speckled</td>
<td>37</td>
</tr>
<tr>
<td>Cultivation status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breeding line</td>
<td>65</td>
<td>Small red</td>
<td>50</td>
</tr>
<tr>
<td>Landrace</td>
<td>201</td>
<td>Red mottled</td>
<td>60</td>
</tr>
<tr>
<td>Variety</td>
<td>130</td>
<td>Red speckled</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yellow</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yellow brown</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 1. STRUCTURE results with 392 single-nucleotide polymorphisms (SNPs). ΔK is given as a function of k (number of clusters). The cluster with the largest ΔK (k = 2) was used to define the number of subpopulations in the Andean diversity panel of common bean.
tend to cluster into separate nodes. There are also separation based on seed type: notably, North American kidney beans form two separate groups and the North American cranberry beans form a group that is separate from the African cranberry beans. The South American accessions are spread throughout the tree and are interspersed with the African materials. The South American ADP lines include the CIAT core collection (11 accessions), the CIAT breeding lines and varieties (seven accessions), and the Ecuador National Program lines (eight accessions).

The genetic diversity in the ADP was analyzed by grouping the germplasm based on gene pool membership into two groups: Andean and Mesoamerican. The Andean group had the highest major allele frequency and the lowest gene diversity (0.11). The Andean–Mesoamerican introgression group (admixed) had the lowest major allele frequency and the highest gene diversity (0.30). The Mesoamerican group had a gene diversity of 0.23 (Table 3). Comparisons within the Andean and Admixed groups by cultivation status revealed that landraces, breeding
lines, and varieties had similar levels of diversity, but the highest gene diversity was in the varieties. There were differences in diversity based on origin. Germplasm from the Caribbean in the ADP was the most diverse and germplasm from Africa was the least diverse (Table 3).

**Genome-Wide Association for Determinacy**

Association mapping for determinacy was conducted to evaluate the utility of the ADP in combination with the SNP genotypic data from the BARCBean6K_3 BeadChip for GWAS. The growth habit of the ADP accessions was characterized as either determinate or indeterminate. A significant region for determinacy was found on chromosome 1 (Fig. 5). The most significant SNP (ss715639272) is 3.1 kb upstream of the TFL1y gene (Phvul.001G189200), which has been identified as the Fin (determinacy) gene (Kwak et al., 2008) (Repinski et al., 2012). This is the closest SNP on the BARCBean6K_3 chip to the TFL gene based on the P. vulgaris genome sequence (Schmutz et al., 2014).

![Figure 4. Neighbor-joining (NJ) tree of Andean accessions of the Andean diversity panel of common bean. The nodes containing the Mesoamerican and introgression lines were collapsed. The NJ tree was developed in TASSEL with 3385 SNP markers using simple parsimony. The numbers 1–11 correspond to genotype groups. Information on individual genotypes is available in Supplemental Table 1 in the order that they appear in the NJ tree. Group 1 includes the Mesoamerican genotypes, which are not included in the figure.](image)

**Table 3. Measures of diversity within Andean diversity panel (ADP) subgroups of common bean, including gene pool, cultivation status, origin, and growth habit.**

<table>
<thead>
<tr>
<th></th>
<th>Major allele frequency</th>
<th>Gene diversity</th>
<th>Heterozygosity</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultivation status (within Andean or admixture)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breeding line</td>
<td>0.908</td>
<td>0.137</td>
<td>0.0064</td>
<td>0.1152</td>
</tr>
<tr>
<td>Landrace</td>
<td>0.917</td>
<td>0.128</td>
<td>0.0067</td>
<td>0.1098</td>
</tr>
<tr>
<td>Variety</td>
<td>0.900</td>
<td>0.150</td>
<td>0.0064</td>
<td>0.1273</td>
</tr>
<tr>
<td>Origin (within Andean or admixture)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>0.923</td>
<td>0.117</td>
<td>0.0056</td>
<td>0.0997</td>
</tr>
<tr>
<td>Caribbean</td>
<td>0.904</td>
<td>0.144</td>
<td>0.0038</td>
<td>0.1219</td>
</tr>
<tr>
<td>North America</td>
<td>0.910</td>
<td>0.132</td>
<td>0.0068</td>
<td>0.1091</td>
</tr>
<tr>
<td>South America</td>
<td>0.910</td>
<td>0.138</td>
<td>0.0034</td>
<td>0.1186</td>
</tr>
<tr>
<td>Growth habit (within Andean or admixture)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determinate</td>
<td>0.912</td>
<td>0.132</td>
<td>0.0074</td>
<td>0.1125</td>
</tr>
<tr>
<td>Indeterminate</td>
<td>0.919</td>
<td>0.123</td>
<td>0.0055</td>
<td>0.1059</td>
</tr>
<tr>
<td>Comparisons by gene pool</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andean</td>
<td>0.924</td>
<td>0.114</td>
<td>0.0049</td>
<td>0.0971</td>
</tr>
<tr>
<td>Mesoamerican</td>
<td>0.825</td>
<td>0.233</td>
<td>0.0053</td>
<td>0.1869</td>
</tr>
<tr>
<td>Admixture</td>
<td>0.792</td>
<td>0.295</td>
<td>0.0307</td>
<td>0.2413</td>
</tr>
<tr>
<td><strong>ADP 396 all</strong></td>
<td>0.887</td>
<td>0.178</td>
<td>0.0066</td>
<td>0.1535</td>
</tr>
</tbody>
</table>

† Admixture includes any accessions that are 10 to 90% Mesoamerican according to the STRUCTURE results.

‡ PIC, polymorphism information content.
Phenotypic Diversity for Economically Important Traits

Subsets of the ADP were evaluated in six locations (three in Africa, one in the Caribbean, and two in the United States) for traits of importance to farmers and consumers to test its utility as a source of germplasm for Andean bean improvement. For each trait, significant phenotypic variability was observed within the ADP. The ADP nursery planted in Zambia in 2014 was under heavy natural CBB pressure. The susceptible check ADP0676 (CELRK) had a score of 7 and the resistant check ADP0665 (USWK-CBB-17) had a score of 1.5. In this trial, 35 lines showed resistance to the pathogen (Fig. 6A). In a field trial in South Africa in 2014, uniform and severe ALS and rust disease pressure were present, as determined from a series of differential cultivars known for their reaction to the diseases planted with the trials. The resistant check, ‘Mexico 54’, had an ALS and rust score of 1, whereas the susceptible check, ‘Montcalm’, had an ALS score of 5.7 and a rust score of 5.3. For both diseases, there was a proportion of the ADP that showed resistance to these diseases (Fig. 6B, C). A set of ADP lines was evaluated for root rot and low fertility resistance in a field in Puerto Rico. The average seed yield of the ADP in this trial was only 100 kg ha\(^{-1}\); in contrast, the Mesoamerican local cultivar check, ‘Verano’ (Beaver et al., 2008), yielded 516 kg ha\(^{-1}\) and the Andean local cultivar check, ‘Badillo’ (Beaver et al., 2010), yielded 645 kg ha\(^{-1}\). This indicates a lack of combined resistance to root rot and low fertility in the panel and a breeding need in this area.

Low soil fertility is a major concern for bean production in East Africa. A field experiment on low P and low N soil in Morogoro, Tanzania, revealed large variability for seed yield under these conditions. Seven ADP lines had yields greater than 1000 kg ha\(^{-1}\), where the mean yield was just 460 kg ha\(^{-1}\) (Fig. 6E). The ADP has some accessions with terminal drought tolerance as shown in a trial in Othello, WA in 2013 where yields ranged from 1100 to 4600 kg ha\(^{-1}\), with an average yield of 2975 kg ha\(^{-1}\) (Fig. 6F). Consumer acceptance and nutritional quality traits have also been evaluated on the ADP. Cooking time was measured in lines grown in Entrican, MI, in 2012 and 2013. Five lines were identified that cooked in less than 27 min (Fig. 6G). Iron concentration was also measured on raw seed of a trial from Othello, WA. Lines were identified with Fe levels as high as 94 µg g\(^{-1}\) (Fig. 6H). Iron bioavailability was measured on the cooked samples and a few lines were identified as having high Fe bioavailability (Fig. 6I).

**DISCUSSION**

The genetic improvement of Andean beans has lagged behind Mesoamerican bean improvement. This may be due, in part, to the narrower genetic diversity of Andean beans, the difficulty in crossing between the two gene pools, and the greater breeding effort focused on Middle American bean improvement (Shii et al., 1980; Sonnante et al., 1994). Andean beans remain an important food source in Africa, Europe, and the Americas. A greater effort in Andean bean breeding and a global coordinated
approach for Andean bean improvement has the potential to result in significant yield gains in Andean beans. The ADP was developed to address this need to improve the rate of improvement in breeding Andean beans for both low-input and high-input production environments.

The characterization of genetic diversity within the ADP was an important step to assess the utility of the panel. The major subpopulations within the ADP are the Andean and the Mesoamerican groups. Some Mesoamerican lines were included in this panel for two reasons: first, there is uncertainty about the gene pool of origin, especially in landraces from Africa; second, there is an overlap of seed types between the Andean and the Mesoamerican gene pools. This is especially true for the red mottled and the yellow market classes. These seed types were most represented in the Mesoamerican and admixed pools and may serve as a bridge to transfer useful traits from Mesoamerican to Andean beans.

As found in previous diversity studies, the Mesoamerican group showed greater diversity than the Andean

Figure 6. Histograms of phenotypic evaluations with subsets of the Andean diversity panel of common bean: (A) Common bacterial blight in Lusaka, Zambia, in 2014; (B) angular leaf spot in Cedara, South Africa, in 2014; (C) rust in Cedara, South Africa, in 2014; (D) seed yield at harvest in root rot pathogen infested soil in Puerto Rico in 2013; (E) seed yield in low fertility soil in Morogoro, Tanzania, in 2013; (F) seed yield under drought stress in Othello, WA, in 2012; (G) cooking time in Entrican, MI, in 2012 and 2013; (H) seed Fe concentration in Othello, WA, in 2012; and (I) seed iron bioavailability in Entrican, MI, in 2012. Resistant (R) and susceptible (S) check values are indicated with arrows.
group (Bellucci et al., 2014; McClean et al., 2012; Sonnante et al., 1994). The admixed group was the most diverse set of accessions in the panel. A number of these lines are varieties and have commercial seed types. Some of these lines have promising disease resistance (e.g., to CBB and ALS) and abiotic stress tolerance, including heat tolerance, and offer potential as parents, especially within the kidney and cranberry market classes.

The Andean ADP lines did not separate into the three races (Nueva Granada, Chile, and Peru). It was supposed that most of the Andean ADP lines would be from Nueva Granada race, since this seed type is most popular in Africa and North America. However, it would have been expected that some of the cranberry beans would have been from the Chile race but there is no clear separation of this group, perhaps because too few accessions in the ADP were from this race or because the race was diluted by hybridization with the Nueva Granada race, either of which would have obscured the difference between the accessions from these races. Studies in other crops have found that simple sequence repeats markers are more efficient than SNP markers for diversity studies (Singh et al., 2013) and this may also influence the ability to separate the Andean ADP into individual races. Other molecular marker–based diversity studies in P. vulgaris have observed a single Andean population (Beebe et al., 2001; Bitocchi et al., 2013).

The PCA shows a tight cluster of North American kidney bean varieties. This possibly represents the stringent selection for industry quality for North American kidney beans. One of the varieties in the group, ‘Red Hawk’, released in 1998 (Kelly et al., 1998), is still a major dark red kidney bean grown today. It meets the strict industry standards for canning quality. As a group, the North American kidney beans (gene diversity = 0.136) are slightly more diverse than North America. As a group, the Andean ADP lines did not separate into the three races (Nueva Granada, Chile, and Peru). It was supposed that most of the Andean ADP lines would be from Nueva Granada race, since this seed type is most popular in Africa and North America. However, it would have been expected that some of the cranberry beans would have been from the Chile race but there is no clear separation of this group, perhaps because too few accessions in the ADP were from this race or because the race was diluted by hybridization with the Nueva Granada race, either of which would have obscured the difference between the accessions from these races. Studies in other crops have found that simple sequence repeats markers are more efficient than SNP markers for diversity studies (Singh et al., 2013) and this may also influence the ability to separate the Andean ADP into individual races. Other molecular marker–based diversity studies in P. vulgaris have observed a single Andean population (Beebe et al., 2001; Bitocchi et al., 2013).

The ADP is a resource for Andean bean improvement. A total of 89 ADP entries were identified as being resistant to the endemic strains of ALS and 84 to rust in the tested locations.

Results from a root rot trial in Puerto Rico indicate that there is limited Andean germplasm in the ADP that shows resistance to the root rot complex present in the Isabela Nursery. Several of the major diseases present in the nursery, including Fusarium solani and Macrophomina phaseolina, are widespread root rot pathogens in African production areas and F. solani is the major root rot pathogen worldwide. Considerable genetic progress is necessary to develop root rot resistance, since almost all commercial Andean cultivars are susceptible to root rot and because large-seeded types tend to be more susceptible to root rot than small-seeded market classes (Abawi and Pastor-Corrales, 1990).

The ADP showed a twofold range in seed Fe concentration in the reported field trial (Fig. 6H), which is similar to other diversity studies in bean (Beebe et al., 2000; Blair et al., 2009). This level of diversity was adequate to identify quantitative trait loci for seed Fe concentration and seed cotyledon Fe concentration in bean (Blair et al., 2009, 2013). Thus, the ADP should provide useful material for assessing concentration differences and the genetic determinants of seed mineral concentrations (Fe and other nutritionally significant minerals) when grown in various control, low fertility, or other abiotic or biotic stress conditions. In addition, this is the first screening of diverse Andean germplasm for Fe bioavailability and the results are promising, since several superior accessions were identified. Notably, a few yellow-seeded landraces from Angola were especially high in Fe bioavailability (Fig. 6I).

The ADP is a resource for Andean bean improvement for diverse production systems in the Americas and Africa. The genotypic and broad phenotypic characterization of the panel can be applied to molecular and phenotypic breeding activities.

Acknowledgments

This work was supported, in part, by funding from the Norman Borlaug Commemorative Research Initiative (US Agency for International Development), and by the USDA Agricultural Research Service through Cooperative Agreement Number 58–6250–0–008 (MAG). The contents of this publication do not necessarily reflect the views or policies of the USDA, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government. We also thank Jose Luis Claros for DNA extraction.
References


