UC Davis

UC Davis Previously Published Works

Title

Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity?

Permalink

https://escholarship.org/uc/item/96m714xn

Journal

Crop Science, 60(3)

ISSN

0011-183X

Authors

Kuzay, Saarah
Hamilton-Conaty, Paige
Palkovic, Antonia
et al.

Publication Date

2020-05-01

DOI

10.1002/csc2.20032

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, availalbe at https://creativecommons.org/licenses/by-nc-nd/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Published online: 9 April 2020

DOI: 10.1002/csc2.20032

Crop Science

ORIGINAL RESEARCH ARTICLE

Plant Genetic Resources

Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity?

Saarah Kuzay | **Paige Hamilton-Conaty** | **Antonia Palkovic** | **Paul Gepts**

Department of Plant Sciences/MS1, University of California, 1 Shields Avenue, Davis, CA 95616–8780, USA

Correspondence
Paul Gepts, Department of Plant Sciences/MS1, University of California, 1 Shields Avenue, Davis, CA 95616–8780, USA.
Email: plgepts@ucdavis.edu

Assigned to Associate Editor Ali M. Missaoui.

Abstract

Core collections are envisioned to be a representative subset of larger germplasm collections. They were introduced to facilitate the characterization and use of these germplasm collections. The common bean (*Phaseolus vulgaris* L.) core collection of the USDA Western Regional Plant Introduction Station was one of the first collections to be established in the early 1990s. Here, we evaluate the representativity of this common bean collection in light of the availability of a single nucleotide polymorphism (SNP) platform and new information about genetic diversity of the species, including phaseolin and seed type data. The SNP diversity was studied with a combination of STRUCTURE, principal coordinate analysis (PCoA), and neighbor-joining analysis (NJA). STRUCTURE analyses were conducted for K (number of subpopulations) = 3 and $K = 7$, based on the ad hoc statistic ΔK. The $K = 3$ analysis recognized the split between Andean and Mesoamerican domesticates and the subdivision of the Mesoamerican domesticates into high- (Durango/Jalisco) and low-altitude (Mesoamerica) ecogeographic races. The $K = 7$ analysis further subdivided the Andean group identified for $K = 3$, as well as the high-altitude group from the Mesoamerican gene pool. It also identified smaller groups consisting of Mesoamerican wild beans. The PCoA and NJA confirmed the STRUCTURE results and highlighted the existence of presumed hybridization among groups. Our results suggest that this core collection should be updated by adding domesticated categories, developing a separate wild common bean core collection, and developing cores for specific purposes.

1 | INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is a grain legume that is consumed principally for its dry seeds or green pods. It plays an important role in human nutrition, mainly in Latin America and Eastern Africa. Its seeds are rich in proteins, fibers, certain minerals like Fe and Zn, and vitamins like folate. Some of these traits, such as the essential amino acid profile, provide complementation with the nutritional characteristics of cereals (Bressani, 1993; Broughton et al., 2003). In addition to their nutritional qualities, common bean, like other food legumes, fixes N symbiotically and thus complements other crops as a source of N in cropping systems (Ehrmann & Ritz, 2014; Hardarson et al., 1993).

Abbreviations: CIAT, Centro Internacional de Agricultura Tropical; GRU, Genetic Resources Unit; NJA, neighbor-joining analysis; PCoA, principal coordinate analysis; SNP, single nucleotide polymorphism; SSR, simple sequence repeat.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2019 The Authors. *Crop Science* published by Wiley Periodicals, Inc. on behalf of Crop Science Society of America

An essential condition for the efficient conservation and use of crop genetic resources is a better understanding of the organization or "structure" of genetic diversity of the crop species, including its wild and domesticated forms (Gepts, 2000, 2014a, 2014b; Hufford, Berny Mier y Teran, & Gepts, 2019). In common bean, a wide range of molecular markers has been used to characterize its genetic diversity, including phaseolin seed protein (Gepts & Bliss, 1985; Gepts, Osborn, Rashka, & Bliss, 1986), allozymes (Koenig & Gepts, 1989; Singh, Nodari, & Gepts, 1991), restriction fragment length polymorphism (RFLP) (Becerra Velasquez & Gepts, 1994), randomly amplified polymorphic DNA (RAPD) (Freyre, Ríos, Guzmán, Debouck, & Gepts, 1996), amplified fragment length polymorphism (AFLP) (Papa et al., 2007; Papa, Acosta, Delgado-Salinas, & Gepts, 2005; Tohme, González, Beebe, & Duque, 1996), simple sequence repeat (SSR) (Blair, Soler, & Cortés, 2012; Kwak & Gepts, 2009), single nucleotide polymorphism (SNP) (Ariani, Berny Mier y Teran, & Gepts, 2016, 2018; Hyten et al., 2010), and DNA sequences (Bitocchi et al., 2012; Cortés, Chavarro, Madrinan, This, & Blair, 2012; Cortés, This, Chavarro, Madrinan, & Blair, 2012; Kami, Becerra Velásquez, Debouck, & Gepts, 1995; Mamidi et al., 2013; Nanni et al., 2011; Rendón-Anaya et al., 2017; Schmutz et al., 2014).

These different approaches have revealed essential features of the organization of genetic diversity (Figure 1). First, the primary gene pool of wild common bean consists of three geographically and genetically diverged gene pools: Mesoamerica (Mexico, Central America, the Andes of Colombia, and Venezuela) vs. central Andes (northern Peru and Ecuador) vs. southern Andes (Argentina, Bolivia, and southern Peru) (Debouck, Toro, Paredes, Johnson, & Gepts, 1993; Freyre et al., 1996; Koenig & Gepts, 1989; Kwak & Gepts, 2009; Rendón-Anaya et al., 2017). The wild gene pool in the central Andes has since been reclassified from *P. vulgaris* to *P. debouckii* A. Delgado based on the age of its divergence, its distinct metabolomic profile, and the existence of ancestral sequences (Figure 1; Ariani, Berny Mier y Teran, & Gepts, 2018; Kami et al., 1995; Rendón-Anaya, Herrera-Estrella, Gepts, & Delgado-Salinas, 2017). The wild gene pools in the central Andes and southern Andes resulted from long-distance dispersal events some 500,000 and 100,000 yr ago, respectively (Ariani et al., 2018). Culturally independent domestications in Mesoamerica and the southern Andes from these diverged wild populations led to distinct domesticated gene pools (Gepts et al., 1986; Kwak & Gepts, 2009; Kwak, Kami, & Gepts, 2009). Second, the current levels of genetic diversity reflect past events before, during, and after domestication, with lower diversity in the two Andean wild gene pools compared with the Mesoamerican one and lesser diversity in the domesticated gene pools compared with their respective ancestral ones (Ariani et al., 2018; Bitocchi et al., 2013; Gepts et al., 1986; Mamidi et al., 2013). Third, further divergence within the two major domesticated gene pools has led to the development of ecogeographic races, three each in the Andean and Mesoamerican gene pools (Singh, Gepts, & Debouck, 1991), with a potential fourth in the Mesoamerican gene pool (Beebe et al., 2000; Díaz & Blair, 2006). The Mesoamerican and Andean gene pools in their centers of domestication are subject to gene flow with their respective wild ancestral forms, which results in feral or weedy local populations. The gene flow, although detectable with molecular markers, does not overcome selection in wild populations; thus, the wild phenotype is maintained in these populations (Beebe, Toro, González, Chacón, & Debouck, 1997; Papa & Gepts, 2003; Papa et al., 2005, 2007). Two of the domesticated ecogeographic races (Jalisco from the Mesoamerican gene pool and Peru from the Andean gene pool) have limited representation in temperate areas, presumably because of unfavorable photoperiod × temperature interactions that delay flowering (Figure 1; e.g., Zeven, 1997; Zeven, Waninge, van Hintum, & Singh, 1999).

Phaseolus genetic resources are conserved in many collections around the world. In addition to the world collection housed at the Centro Internacional de Agricultura Tropical (CIAT; Cali, Colombia), which counts some 38,000 accessions (https://ciat.cgiar.org/what-we-do/crop-conservation-and-use/bean-diversity/ [accessed 4 Feb. 2019]), one of the largest collections is held by the USDA Western Regional Plant Introduction Station at Pullman, WA, USA, with some 17,940 accessions of *Phaseolus* sp. and 14,075 of *P. vulgaris* (https://npgsweb.ars-grin.gov/gringlobal/search.aspx [accessed 4 Feb. 2019]). To address the logistical difficulties associated with the characterization and evaluation of collections of such large size, the concept of core collection was introduced by O. Frankel and further developed by A. H. D. Brown (Brown, 1989; Brown & Spillane, 1999; Hodgkin, Brown, van Hintum, & Morales, 1995). Core collections consist of 5–10% of the collection at large and are representative—qualitatively and quantitatively—of this collection. They are meant to constitute an entryway to the larger collection; variations of the core collection approach have been proposed such as collection focused on a specific region (e.g., Chile; Paredes, Becerra Velásquez, Tay Urbina, Blair, & Bascur, 2010), trait (e.g., determinacy; Kwak, Toro, Debouck, & Gepts, 2012), type of material (e.g., wild vs. domesticated; Tohme et al., 1996), or mini-core collections (Brown & Spillane, 1999; Gepts, 2006; Upadhyaya, Wang, Gowda, & Sharma, 2013).

In this article, we present the results of an analysis of the genetic diversity of a USDA core collection of common bean with a SNP chip platform of some 5400 markers distributed over the 11 chromosomes (Pv01–Pv11) of the species (Song et al., 2015). This collection was previously analyzed with SSR markers (McClean et al., 2012). In the discussion, we compare the results of these two studies and examine the

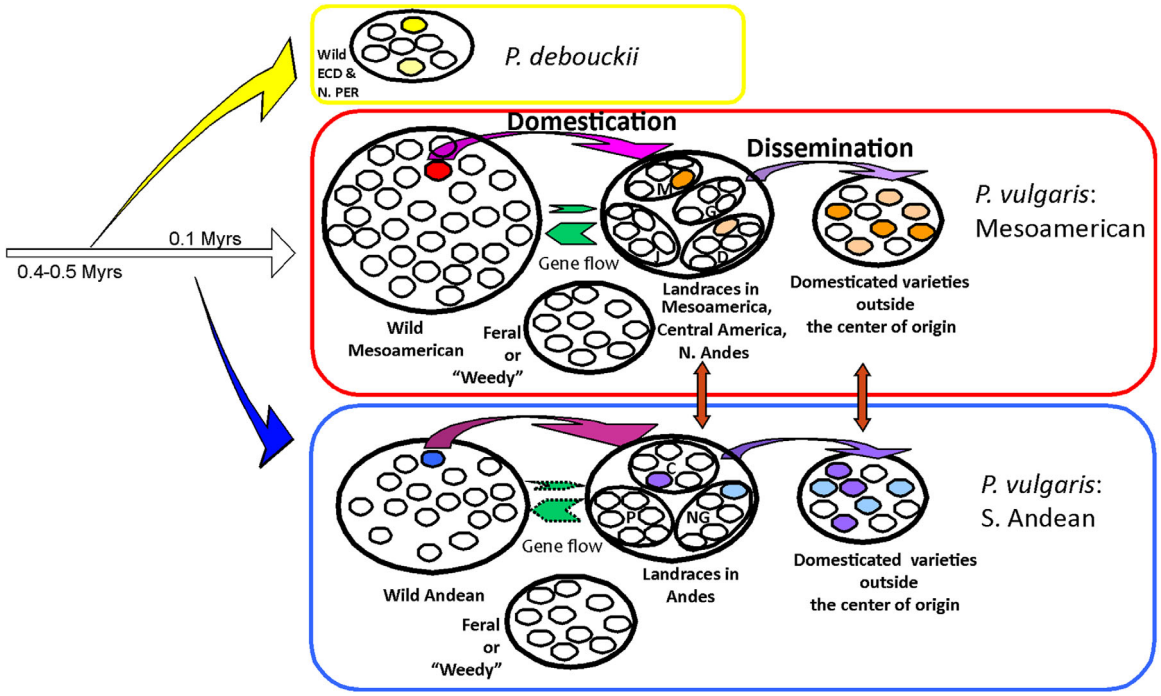

FIGURE 1 General representation of the wild and domesticated gene pools of common bean (*Phaseolus vulgaris*) and its wild sibling species *P. debouckii*. The ellipses in bold represent major geographic wild or domesticated gene pools. The smaller ellipses within these gene pools represent individual populations. For further explanations, see text. Myrs, million years; ECD, Ecuador; N. PER, northern Peru. Ecogeographic races: C, Chile; D, Durango; G, Guatemala; J, Jalisco; M, Mesoamerica; NG, Nueva Granada; P, Peru

representativity of this core collection in light of the several studies cited above that have examined the genetic diversity of wild and domesticated common bean.

2 | MATERIALS AND METHODS

2.1 | Plant material

The current core collection for common bean held at the Western Regional Plant Introduction Station of the USDA in Pullman, WA, is an amalgamation of an original collection focused mainly on Mesoamerican gene pool accessions (n = 224) with a subsequent addition of Central (n = 101) and South American (n = 97) accessions for a total of n = 422 (USDA-ARS, National Genetic Resources Program, Germplasm Resources Information Network [GRIN], [Online Database] National Germplasm Resources Laboratory, Beltsville, MD. Available: https://www.ars-grin.gov/npgs/cgc_reports/phascgc.htm [accessed 23 Nov. 2019]). Plants were grown in the greenhouse to produce leaf tissue for DNA extraction. One plant per accession was analyzed because the emphasis was on determining genetic diversity patterns across broad groups of accessions, rather than within accessions. For various reasons, including photoperiod sensitivity and other causes of lack of adaptation to University of California, Davis, greenhouse conditions, only 363 entries of the core collection could be studied. A subset of these 363 accessions is also maintained at the Genetic Resources Unit (GRU) at CIAT (Cali, Colombia). The database maintained by the GRU was used to retrieve, when available, seed photos and information on the type of phaseolin seed protein of share accessions (https://genebank.ciat.cgiar.org/genebank/language.do?collection = bean).

Each USDA accession was planted in the greenhouse and tissue from the first trifoliate leaf was harvested, freeze dried, and ground into powder via bead beating. DNA from each accession was extracted with QIAGEN DNeasy Plant Mini Kit and quantified with the 260/280 and 260/230 absorbance ratios (optimal ratios of 1.7–1.9 and 2.0–2.2), respectively, with the NanoDrop. To assess overall quality of the genomic DNA, 5 µl of DNA was run using standard gel electrophoresis. Only high-quality genomic DNA with concentrations > 100 ng μl^{-1} and with no evidence of DNA degradation or protein or RNA contamination was used for SNP genotyping.

2.2 | SNP genotyping

Bean DNA samples were genotyped with the Illumina BARCBean6K_3 Infinium SNP array (Song et al., 2015),

which allows 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. The SNP calling was conducted with the genotyping module V2011.1 of the GenomeStudio software (Illumina). The SNP dataset is publicly available, as it has been deposited in the University of California, Davis, Library Dash database of the California Digital Library (https://doi.org/10.25338/B8KP45) (Gepts, Kuzay, & Hamilton, 2019).

2.3 | Genetic diversity analyses

To distinguish the different subpopulations within the 363 accessions, an analysis of the SNP data was performed using the programs STRUCTURE (Falush, Stephens, & Pritchard, 2003; Pritchard, Stephens, & Donnelly, 2000) and GenAlEx (Peakall & Smouse, 2012) to determine population structure and principal coordinate analysis (PCoA), respectively. Per methods in Kwak and Gepts (2009), STRUCTURE was run at five simulations per K value (number of subpopulations) from $K = 2$ to $K = 7$ using the admixture model with 500 replicates for burn-in and 5000 replicates during the analysis. STRUCTURE Harvester (Earl & von Holdt, 2012) was then used to identify the optimum K values with the maximum Delta statistical test. For each optimal K value, membership coefficients with the lowest likelihood value were used to assign the accessions to the subpopulations (Kwak & Gepts, 2009). The STRUCTURE plots generated from this analysis were bar graphs developed in Excel 2010. Analysis of the genetic relatedness among different accessions was visualized in a PCoA plot using the GenAlEx program 6.501 (Peakall & Smouse, 2012) program in Excel 2010. Membership in the K number of groups or subpopulations was determined by membership coefficient values above .80. Accessions with multiple membership coefficients below .80 were not assigned to a specific subpopulation. To further resolve separation across the proposed subpopulations, SNP data for all accessions were used to generate a neighbor-joining tree using Molecular Evolutionary Genetics Analysis version 5 (MEGA5) (Tamura et al., 2011). Within MEGA5, a neighbor-joining algorithm uses a matrix of pairwise distances estimated using the Tamura and Nei (1993) model to generate this evolutionary tree.

3 | RESULTS

3.1 | SNP diversity

The BARCBean6K_3 Infinium SNP array includes 5398 functional SNP markers (Song et al., 2015). Because our sample consisted of 363 entries, the total number of data points was 1,959,474. Of these, 780,306 (or 40%) were A alleles and 1,039,690 (or 53%) were B alleles. The number of missing or heterozygous data points was 139,478 (or 7%). Over 80% of SNPs had < 10% missing data (Supplemental Figure S1).

3.2 | STRUCTURE analysis

A preliminary analysis of the data with the Evanno, Regnaut, and Goudet (2005) test showed two peaks in the ΔK value, a major one at $K = 3$ and a minor one at $K = 7$ (Supplemental Figure S2). Hence, we conducted further analysis and interpretation for these two K values.

3.2.1 | Structure analysis for $K = 3$

For $K = 3$ (Table 1), it was possible to determine the general identity of the three populations based on current data and results from previous studies (Pallottini, Garcia, Kami, Barcaccia, & Gepts, 2004; Singh et al., 1991). Two-thirds of Population K3.1 (Figure 2A, $n = 79$) originated primarily from Andean countries and had large seed weights (on average, > 50 g 100 seed^{-1}), based on information available from the database maintained by the GRU at CIAT (Cali, Colombia) (https://genebank.ciat.cgiar.org/genebank/language.do?collection = bean, [accessed 18 Oct. 2014]). This same database also showed that accessions in this group had predominantly an Andean phaseolin protein electrophoretic type (T or C; Gepts et al., 1986). An exception to this pattern was accession G18971, a climbing accession with small, black seeds and an Sd phaseolin type, of Mesoamerican origin (Koenig, Singh, & Gepts, 1990). Consistent with the large seed size and the Andean geographic origin, many of the member accessions of Population K3.1 had elongated seeds characteristic of race Nueva Granada (Singh et al., 1991; Supplemental Figure S3).

Population K3.2 (Figure 2A, $n = 89$) had a predominantly Mesoamerican origin (77 of 89 accessions or 85%). This population was widespread in Mexico and Central America, with more limited representation in the northern Andes. Average seed size was small (25 g 100 seed^{-1}) in Mesoamerica and Colombia. Aside of one accession in Ecuador, no other Andean countries were represented in this population. Most accessions for which this data were available from the CIAT GRU database showed an "S" or "Sd" phaseolin type, of Mesoamerican origin. Three accessions from Nicaragua (G2070, G2078, and G2084) showed a "CH" phaseolin type, originating among Colombian wild beans (Gepts & Bliss 1986). The majority of accessions had dark red or matte black seeds, with the exception of a few accessions with lighter-colored seeds (Supplemental Figure S3).

TABLE 1 Geographic origins and seed weight distribution in subpopulations identified by STRUCTURE at K = 3

	K3.1		K3.2		K3.3		K3.Other	
Country	***N***	**SeedW[a]**	***N***	**SeedW**	***N***	**SeedW**	***N***	**SeedW**
Andean geographic area								
Bolivia	5	49	0		0		0	
Peru	18	48	0		0		3	20
Ecuador	16	50	1	49	0		1	30
Colombia	13	59	11	25	4	26	10	40
Total	52	51	12	37	4	26	14	30
Mesoamerican geographic area								
Mexico	14	48	31	21	99	35	45	15
Guatemala	3	70	8	23	7	29	15	24
El Salvador	2	67	8	25	1	27	3	30
Honduras	1	77	7	23	0		2	24
Nicaragua	0		11	24	2	26	0	
Costa Rica	3	55	12	27	0		1	55
Total	23	66	77	25	109	29	66	30
General total	75	59	89	31	113	28	80	30

[a]SeedW, seed weight.

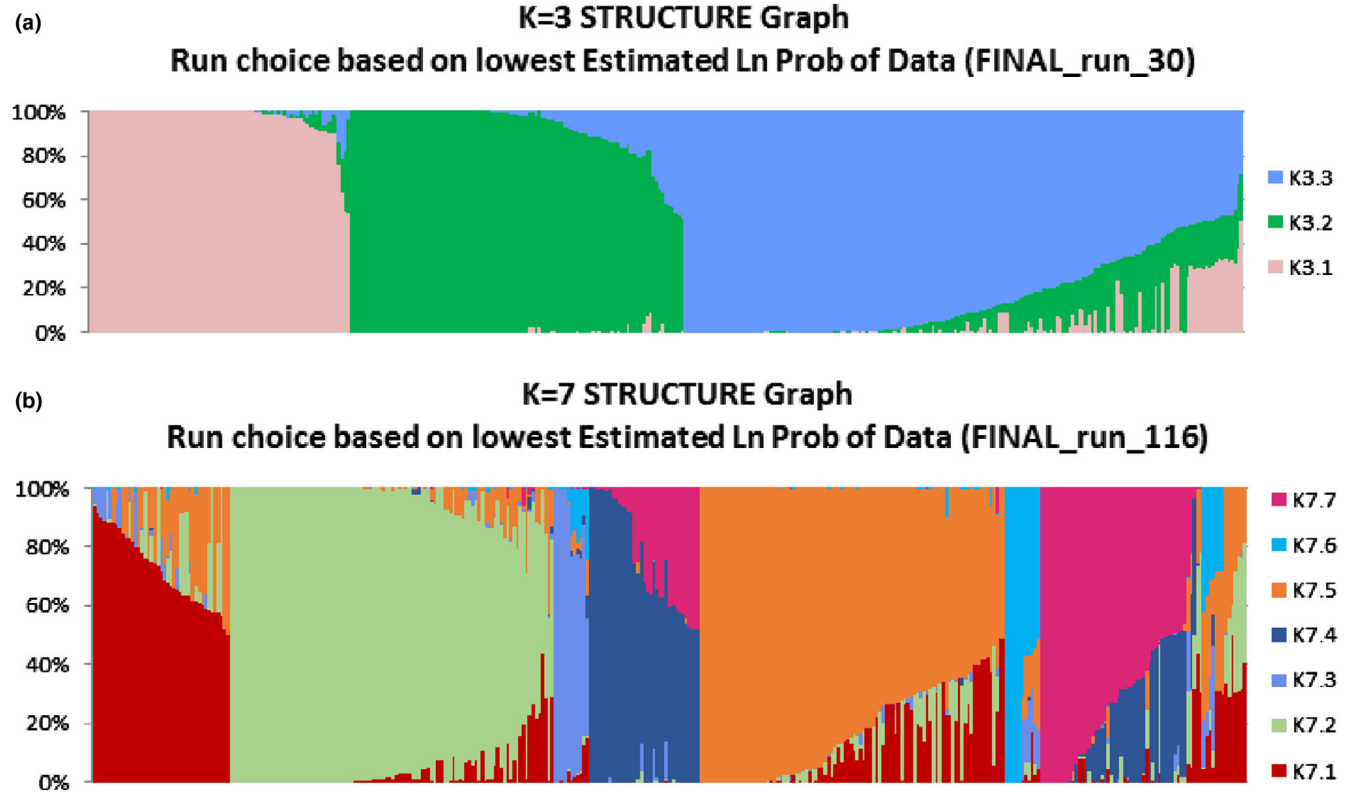

FIGURE 2 STRUCTURE analysis of genetic diversity of the USDA *Phaseolus vulgaris* core collection. (A) $K = 3$. (B) $K = 7$

In Population K3.3 (Figure 2A), close to 90% of accessions came from Mexico. The rest came from Central American countries and Colombia. Seed size was small (< 30 g 100 seed^{-1}), except in Mexico where seeds were medium sized (average of 35 g 100 seed^{-1}). All accessions for which this data type was available showed a Mesoamerican phaseolin type, mostly "S," but also "Sb" and "Sd" (Koenig et al., 1990). Two accessions showed a phaseolin type normally observed

in wild types, G878 ("M1," from Hidalgo state) and G2494 ("M," from the state of Oaxaca). Striking in this population was the higher frequency of accessions with tan- or yellow-colored seeds.

Accessions included in Populations K3.1 to K3.3 all had membership coefficients in their respective populations > 80%. Accessions with lower membership coefficients amounted to some 20% of our total sample. This "hybrid" group was actually quite diverse based on seed weight, seed phenotypes, and phaseolin types (Figure 2A, Supplemental Figure S3, Supplemental Table S1). In particular, it included a group of some 10 wild beans (e.g., Supplemental Figure S3; from 276_PI 318695_G12864 to 304__PI 325691_12883).

3.2.2 | Structure analysis for $K = 7$

In the $K = 7$ STRUCTURE analysis, Population K7.1 included 14 accessions distributed exclusively in Mexico and some Central American countries (Figure 2B, Table 2). There were no accessions from Andean countries. Seeds were small (26 g 100 seed^{-1}) and predominantly black (Supplemental Figure S4). The large ($n = 90$) K7.2 group (Figure 2B) was distributed in Mesoamerica, as well as in the Andes (from Colombia to Peru). In Mesoamerica, seeds were small and comparable in size to the K7.1 group (25 g 100 seed^{-1}), whereas in the Andes, seeds were medium sized (36 g 100 seed^{-1}). Most seed colors were red or black, the latter being non-shiny (Supplemental Figure S4).

The K7.3 group (Figure 2B) was small and consisted of five wild lines from Mexico with an average seed weight of 3 g 100 seed^{-1}. Seed colors included gray mottled on cream background as well as cream-tan, usually observed among wild beans. Group K7.6 (Figure 2B) contained five wild lines from Mexico, with larger seeds (12 g 100 seed^{-1}), but still below the smallest domesticated seed weight (~19–20 g 100 seed^{-1}). Group K7.4 (Figure 2B, $n = 14$) was a predominantly Andean group with representatives from Ecuador and Peru with large seeds (average of 53 g 100 seed^{-1}) (Table 3). Photos of only two accessions were available from the CIAT database, which showed both accessions with roundish seeds and vivid colors, either red or blackish striping over light-cream color (Supplemental Figure S4). In contrast, group K7.5 (Figure 2B) consisted of 52 accessions from Mesoamerica and two from Colombia. The average seed weight was intermediate between small and medium seeded (around 31 g 100 seed^{-1}). It included mostly accessions with seeds that were cream to tan colored. The exceptional black-colored seeds were shiny in contrast with the small-seeded, dull black accessions found in group K7.2 (Supplemental Figure S4).

Group K7.7 (Figure 2B) was distributed predominantly in the Mesoamerican region (16 entries), but it also included four accessions distributed in the Andean region of Colombia, Ecuador, and Peru. In contrast with group K7.5, accessions of this group had large seeds (average of > 47 g 100 seed^{-1}) even in the Mesoamerican region. Seeds were typically elongated of varying color, including red, yellow, or red or purple mottled (Supplemental Figure S4). Finally, the largest group ($n = 155$) was the group with membership coefficients below .80 in any of the previous groups. This group was distributed in the Andean and Mesoamerican regions. Average seed weights were higher in the Andean region (46 g 100 seed^{-1}) than in the Mesoamerican regions (32 g 100 seed^{-1}).

3.3 | Neighbor-joining analysis

The neighbor-joining tree clarified the separation among the different groups identified by the STRUCTURE analysis. For $K = 3$ (Figure 3), a major gap separated K3.1 from K3.2 and K3.3. K3.1 was distributed for two thirds in the Andes but also included representatives from Mesoamerican countries. The groups K3.2 and K3.3 were closer together than either was to K3.1. Furthermore, the group K3.OTHER with membership coefficients below .8 was located in an intermediate position providing some evidence that their low membership coefficient could be due to hybridization among the different groups, either between K3.2 and K3.3, or between these two groups and K3.1. Furthermore, a small group of accessions located in the neighbor-joining tree between group K3.1, on one hand, and groups K3.2 and K3.3, on the other hand, showed more marked divergence than the other K3 groups. K3.Other included two accessions (22 and 24) with small (22 to 23 g 100 seed^{-1}), black seeds, one each from Guatemala and Peru, and a S or Sd phaseolin (Koenig et al., 1990), three accessions from Colombia (58, 59, and 336) with large seeds (45–56 g 100 seed^{-1}) and Andean (Gepts et al., 1986; Koenig et al., 1990) phaseolin types, and one additional accession from Peru (266) with medium-sized seeds (33 g 100 seed^{-1}).

For $K = 7$ (Figure 4), the K3.1 group was further subdivided into groups K7.4, K7.7, and K7.OTHER, whereas group K3.2 remained largely intact in group K7.2. Group K3.3 was subdivided into groups K7.1 and K7.5. The K3.OTHER group remained largely intact, but two smaller groups, K7.3 and K7.6, were separated from the larger group, and in a position that was intermediate between Andean and Mesoamerican groups, but closer to the latter. Each of these contained exclusively wild common bean accessions showing a mix of gray-brown mottled or cream-colored small seeds (Supplemental Figure S4).

3.4 | Principal coordinate analysis

The PCoA provided separation along the first axis (72% of the total variance) into a tightly knit group (corresponding

TABLE 2 Geographic distribution of common bean accessions from the core collection and their respective seed weight averages for $K = 7$ in STRUCTURE

	K7.1		K7.2		K7.3		K7.4		K7.5		K7.6		K7.7		K7.Other[a]	
Country	**N**	**Seed W[b]**	**N**	**Seed W**	**N**	**SeedW**	**N**	**SeedW**	**N**	**SeedW**	**N**	**SeedW**	**N**	**SeedW**	**N**	**SeedW**
Andean countries																
Bolivia	0		0		0		0		0		0		0		5	49
Peru	0		3	42	0		7	41	0		0		2	50	9	45
Ecuador	0		2	39	0		6	64	0		0		1	32	9	42
Colombia	0		11	25	0		0		2	32	0		1	61	26	49
Total Andean	0		16	36	0		13	53	2	32	0		4	47	49	46
Mexico and Central American countries																
Mexico	8	32	29	23	5	3	1	21	50	37	5	12	10	46	81	27
Guatemala	4	28	7	24	0		0		1	26	0		3	25	18	24
El Salvador	0		8	25	0		0		0		0		1	74	4	29
Honduras	1	22	7	23	0		0		0		0		0		1	27
Nicaragua	1	23	11	24	0		0		1	29	0		0		0	
Costa Rica	0		12	27	0		0		0		0		2	55	2	55
Total Meso[c]	14	26	74	25	5	3	1	21	52	31	5	12	16	50	106	32
Total sample	14	26	90	30	5	3	14	37	54	32	5	12	20	49	155	39

[a]K7.Other, group of accessions with STRUCTURE membership coefficient less than .80.
[b]SeedW, seed weight.
[c]Meso, Mesoamerican.

TABLE 3 "Hybridity" of the composition of *K* groups defined by STRUCTURE analysis

	Origin of marker alleles							
***K* group**	**Description**	**K3.1**	**K3.2**	**K3.3**				
A. $K = 3$								
K3.1	Andean; race Nueva Granada	**0.967**[a]	0.021	0.012				
K3.2	Mesoamerican; race Mesoamerica	0.003	**0.917**	0.079				
K3.3	Mesoamerican; races Durango/Jalisco	0.053	0.108	**0.839**				
		K7.1	**K7.2**	**K7.3**	**K7.4**	**K7.5**	**K7.6**	**K7.7**
B. $K = 7$								
K7.1	Mesoamerican; races Durango/Jalisco from Mexico & Central America	**0.721**	0.080	0.031	0.000	0.166	0.003	0.000
K7.2	Mesoamerican; race Mesoamerica	0.038	**0.922**	0.002	0.002	0.034	0.001	0.001
K7.3	Mesoamerican wild; smaller seeded	0.035	0.006	**0.818**	0.015	0.030	0.094	0.001
K7.4	Andean; Ecuador & Peru	0.000	0.002	0.010	**0.752**	0.002	0.000	0.234
K7.5	Mesoamerican; races Durango/Jalisco from Mexico mainly	0.116	0.052	0.004	0.000	**0.825**	0.002	0.001
K7.6	Mesoamerican wild; larger seeded	0.025	0.014	0.089	0.001	0.101	**0.770**	0.000
K7.7	Andean; Mexico & Central America	0.083	0.075	0.018	0.158	0.089	0.035	**0.541**

[a]Bold: Each row represents an estimate of the origin of marker alleles contained in the respective K group. Numbers in bold represent the proportion of alleles from the same group.

to group K3.1), on the one hand, and a more diffuse group (corresponding to groups K3.2 and K3.3), on the other (Figure 5). The groups K.2 and K3.3 were well separated along the second principal coordinate (13% of the variance). The K3.OTHER group, with membership coefficient in any of the STRUCTURE groups below .8, was located within the coordinate space defined by the three STRUCTURE groups but generally did not overlap with these groups.

When the topology of the PCoA was examined in light of the $K = 7$ STRUCTURE analysis, the two groups identified with the $K = 3$ analysis along the second axis were still well defined. However, the tightly knit group along the first axis appeared to be more heterogeneous. The K7.OTHER now not only occupied space in between the other K7 groups, but also showed a significant overlap with these groups.

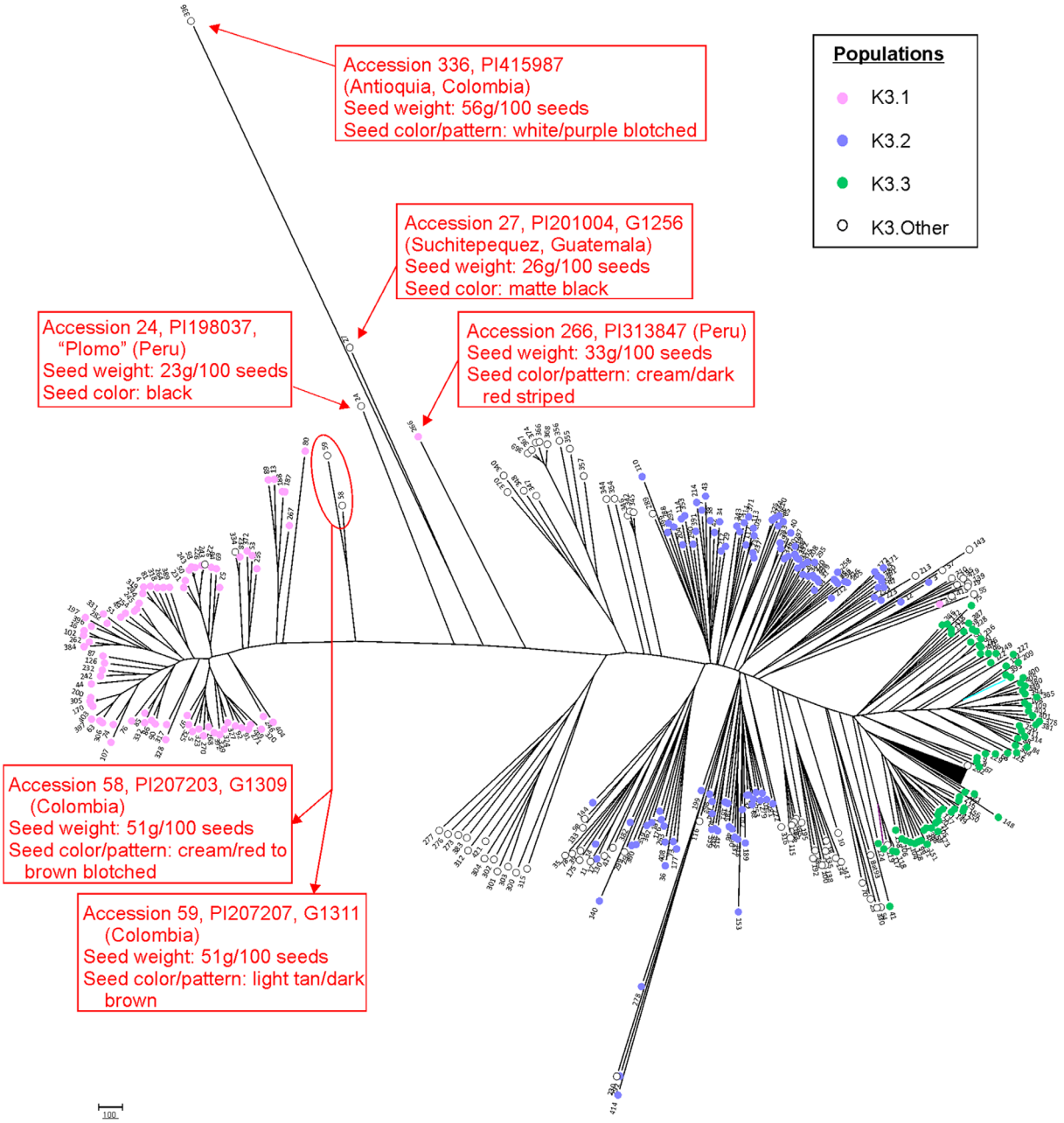

FIGURE 3 Neighbor-joining analysis of single nucleotide polymorphism (SNP) diversity in the USDA core collection of common bean following a STRUCTURE analysis for $K = 3$ with a .8 membership coefficient cutoff. Color coding and K group numbering are as in Figures 2A and 5

4 | DISCUSSION

4.1 | Identity of the groups identified in this study

The identity of groups on the basis of SNP variation with the STRUCTURE software (Pritchard et al., 2000) can be described further using data on geographic distribution, seed type (color, color pattern, size, and shape), and phaseolin seed protein electrophoretic types. The latter two types of data were obtained from the CIAT GRU website.

For $K = 3$, the three groups showed differential geographic distribution. Group K3.1 was of Andean origin, based on its primary distribution in Andean countries, its large seed weight (> 50 g 100 seed^{-1}, Table 1), and the high frequency of Andean electrophoretic types for phaseolin seed protein (Koenig et al., 1990; Supplemental Table S1). The Mesoamerican representatives in this group are probably the result of introduction of Andean types into Mexico and Central America. Such observations have been made before (Gepts et al., 1986). Furthermore, the predominance of elongated seeds (Table 1, Supplemental Figure S3) suggests

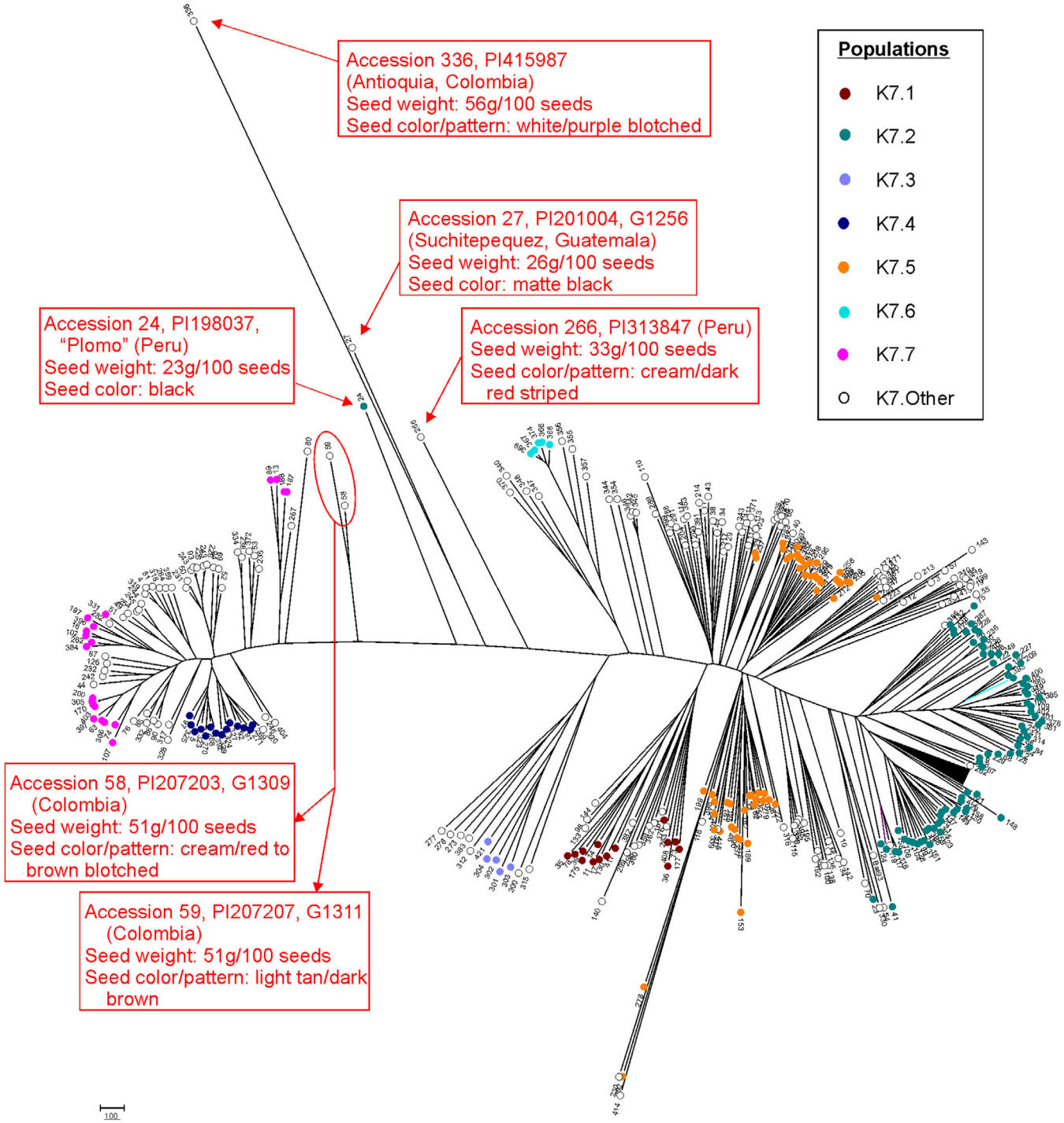

FIGURE 4 Neighbor-joining analysis of single nucleotide polymorphism (SNP) diversity in the USDA core collection of common bean following a STRUCTURE analysis for $K = 7$ with a .8 membership coefficient cutoff. Color coding and K group numbering are as in Figures 2B and 6

that this Andean group largely represents the ecogeographic race Nueva Granada (Singh et al., 1991).

In contrast, groups K3.2 and K3.3 were of Mesoamerican evolutionary origin based on their geographic provenance. Both groups were represented in Mexico and Central America but were also distributed in the northern Andes. K3.2 was more abundantly represented in the Central American countries, as well as Colombia and Ecuador. K3.3 was only present in some of the Central American countries (Guatemala, El Salvador, and Nicaragua) and Colombia, suggesting a distribution focused primarily in Mexico ($n = 99$, Table 1). The relative distribution of the two

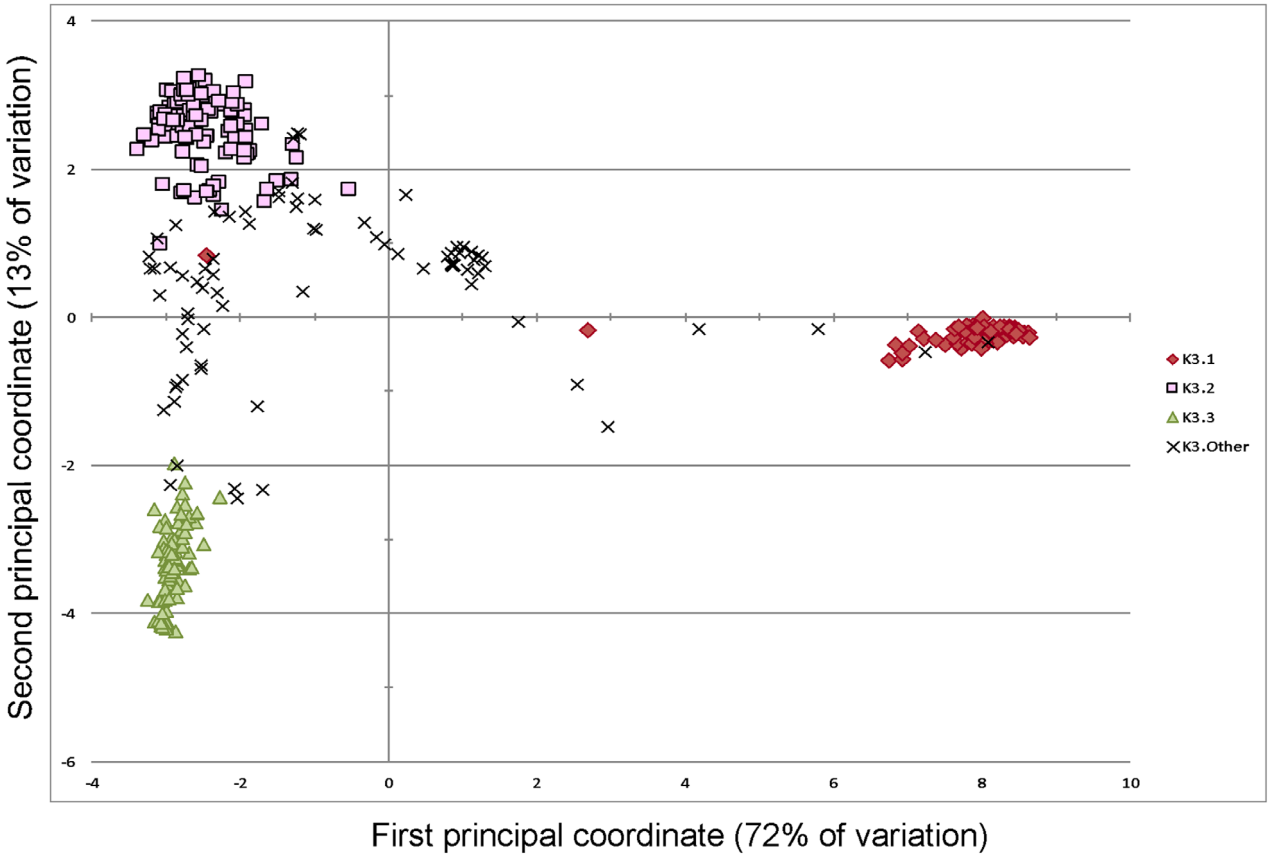

FIGURE 5 Principal coordinate analysis of single nucleotide polymorphism (SNP) diversity in the USDA core collection of common bean. Population assignments are based on $K = 3$ STRUCTURE analysis with .8 membership coefficient cutoff. Color coding and K group numbering are as in Figures 2A and 3

groups in Mexico, Central America, and the northern Andean countries of Colombia and Ecuador, as well as seed type (seed weight [Table 1], seed color [Supplemental Figure S3]) and phaseolin types (Supplemental Table S1), suggest that group K3.2 represents ecogeographic race Mesoamerica, adapted to warmer, more humid lowlands, whereas group K3.3 represents the ecogeographic race Durango, adapted to drier highlands, and potentially also Jalisco, adapted to the more humid highlands. Inspection of seed types in Supplemental Figure S3 suggests, however, reciprocal gene flow between groups K3.2 and K3.3, as might be expected between two conspecific, largely sympatric groups.

Group K3.Other included a mixture of different seed types also seen in the groups discussed above, as well as several wild *P. vulgaris* accessions. Their low membership coefficients (< .80) suggest hybridization among domesticated types or a more distant relationship for the wild beans present in this group.

For $K = 7$, two of the three K3 groups (K3.1 and K3.3) were subdivided into smaller groups, whereas the K3.2 group remained unsubdivided. The Andean group K3.1 was subdivided into two groups: K7.4 and K7.7. Based on seed type (color, shape, and size; Supplemental Figure S4) and phaseolin type (Supplemental Table S2), these two groups were Andean in origin. With one exception, group K7.4 originated in Ecuador and Peru. The seeds were large, especially in Ecuador (Table 1), and round shaped, suggesting they may be representatives of race Peru (Singh et al., 1991; Supplemental Figure S4). In contrast, K7.7, although Andean, was mainly distributed in Mexico and Central America. As stated above, accessions of K7.7 represent introductions of Andean bean varieties in the Mesoamerican bean area. The predominant seed type was large and elongated, characteristic of race Nueva Granada (Singh et al., 1991). The remaining K3.1 members, not included in groups K7.4 and K7.7, were included in group K7.Other with membership coefficients below .80.

Group K3.2 was not subdivided further under $K = 7$ (group K7.2), reflecting the distinctiveness and homogeneity of this group, generally characterized by small, dull black, or red seeds, and indeterminate, upright bush plant types (Singh et al., 1991; Supplemental Figure S4). Group K3.3, under $K = 7$, was further subdivided into two groups, K7.1 and K7.5. The geographic distribution of these two groups differed in that K7.1 was present in Mexico but included a relatively important representation in the northern countries of Central

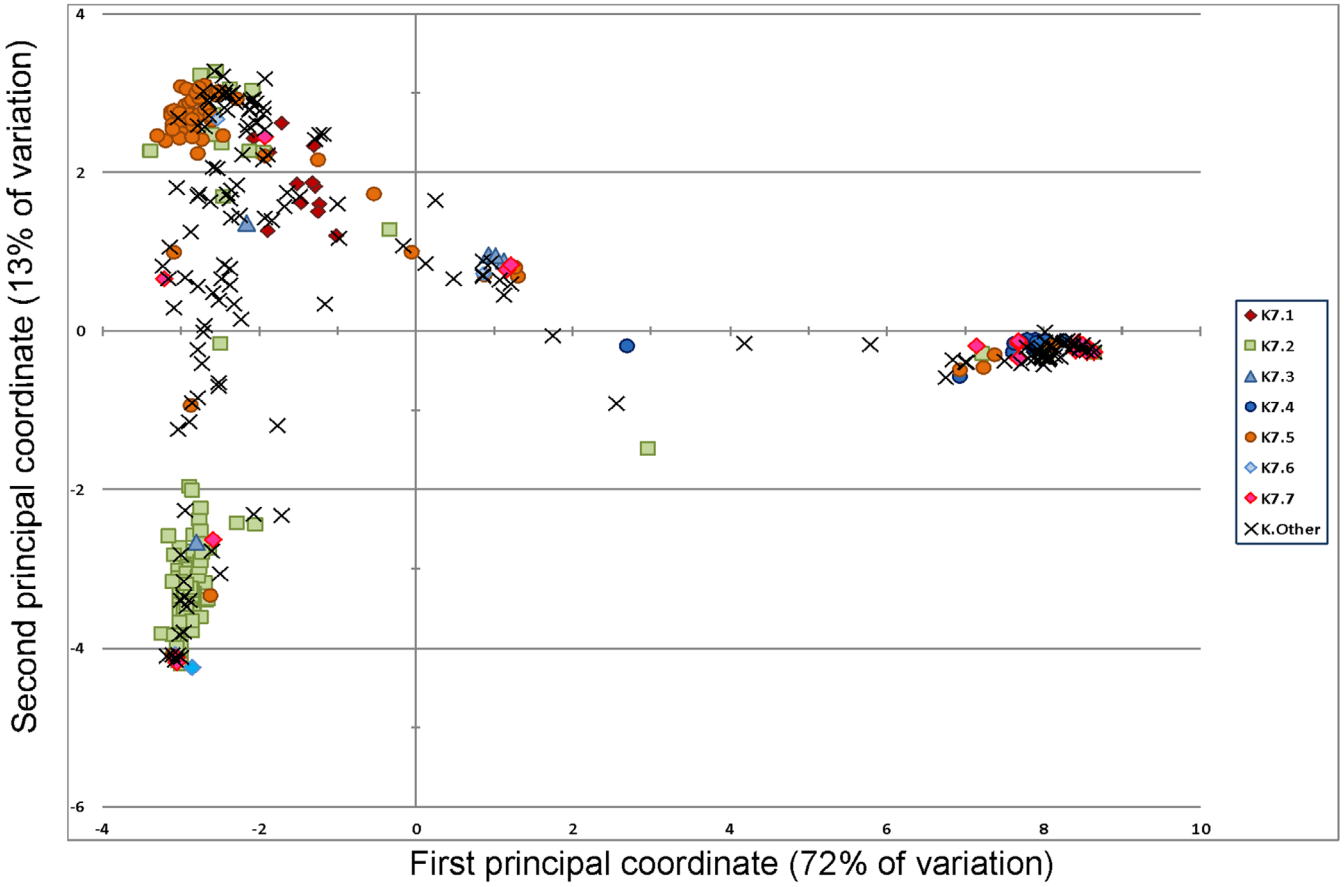

FIGURE 6 Principal coordinate analysis of single nucleotide polymorphism (SNP) diversity in the USDA core collection of common bean. Population assignments are based on $K = 7$ STRUCTURE analysis with .8 membership coefficient cutoff. Color coding and K group numbering are as in Figures 2B and 4

America (Guatemala, Honduras, and Nicaragua). K7.5, on the other hand, was distributed almost entirely in Mexico. Inspection of the seed types of these two groups suggests that K7.5 represents ecogeographic race Durango of the northern, more arid highlands of Mexico. The identity of the K7.1 group is undetermined at this stage. It could include representatives of races Jalisco (Singh et al., 1991) or Guatemala (Beebe et al., 2000; Díaz & Blair, 2006), of the more humid highlands of Mexico and Central America.

4.2 | Comparison between SSR and SNP analyses

McClean et al. (2012) conducted an analysis of SSR diversity in the same core collection. Their sample included 171 entries of the collection and 46 polymorphic SSRs, compared with the 357 lines and 5398 SNPs in the current study. McClean et al. (2012) identified three Andean and six Mesoamerican populations using STRUCTURE (Falush et al., 2003; Pritchard et al., 2000). Two additional groups were defined in the Andean and Mesoamerican gene pools for those accessions whose ancestry was lower than 70% in any of the respective gene pools.

We compared the population subdivisions in the current and McClean et al. (2012) studies were comparable (Tables 4 and 5, for $K = 3$ and $K = 7$, respectively). Different groups identified by McClean et al. (2012) as belonging to both the Andean and Mesoamerican gene pools were classified in this study as being Andean (Table 4, group K3.1). A similar observation can be made for groups K3.2 and K3.3; nevertheless, one can observe that a higher frequency ($n = 14$) of the MA:CA group (Central American entries of the Mesoamerican gene pool) corresponds to the K3.2 group, which represents race Mesoamerica. Similarly, one observes that a higher frequency of the MA:N-Mx (northern Mexico) and MA:SC-Mx (south-central Mexico) ($n = 11$ and $n = 14$, respectively) corresponds to the K3.3 group (races Durango/Jalisco) (Table 4). For the comparison with $K = 7$ subdivisions, the distribution shows a reduced number of Andean materials between $K = 3$ ($n = 17$) and $K = 7$ ($n = 6$), because of a reclassification in the current study of Andean to Mesoamerican materials. Furthermore, there was a general correspondence between the Mesoamerican domesticated groups identified in

TABLE 4 Comparison of groupings between this study ($K = 3$) and those of McClean et al. (2012)

			McClean et al. (2012)													
			Andean[a]				Mesoamerican[b]									
Gene pool	Subdivision (race)	STRUCTURE group	A: N-SA	A: C-SA	A: S-Mx-SA	A: Mixed	MA: N-Mx	MA: NC-Mx	MA: SC-Mx	MA: S-Mx	MA: CA	MA: N-SA	MA: Mixed	Total	Not analyzed by McClean et al. (2012)	Not analyzed in this study
This study																
Andean		3.1	1	5	2	1	1	0	1	0	1	4	1	17	59	
Mesoamerica	Mesoamerica	3.2	0	0	4	0	0	1	1	0	14	3	7	30	59	
	Durango/Jalisco	3.3	1	0	1	1	11	2	14	4	0	0	6	40	75	
		3.Other (< .80)	0	1	0	0	1	5	1	6	3	3	8	28	49	
		Total	2	6	7	2	13	8	17	10	18	10	22	115	242	26
Total analyzed in this study							357									

[a]Groupings of McClean et al. (2012) for the Andean gene pool: A:N-SA, northern South America; A:C-SA, central South America; A:S-Mx-SA, southern Mexico and South America; A:Mixed, mixed.

[b]Groupings of McClean et al. (2012) for the Mesoamerican gene pool: MA:N-Mx, northern Mexico; MA:NC-Mx, north-central Mexico; MA:SC-Mx, south-central Mexico; MA:S-Mx, southern Mexico; MA:CA, Central America; MA:N-SA, northern South America; MA:Mixed, mixed.

TABLE 5 Comparison of groupings between this study ($K = 7$) and those of McClean et al. (2012)

			McClean et al. (2012)													
			Andean[a]				Mesoamerican[b]									
Gene pool	Subdivision or region	STRUCTURE group	A: N-SA	A: C-SA	A: S-Mx-SA	A: Mixed	MA: N-Mx	MA: NC-Mx	MA: SC-Mx	MA: S-Mx	MA: CA	MA: N-SA	MA: Mixed	Total	Not analyzed by McClean et al., 2012	Not analyzed in this study
This study																
Andes	Mexico, Central America	7.7	0	1	1	0	0	0	0	0	1	0	0	3	17	
	Race Peru	7.4	0	2	0	0	0	0	0	0	0	1	0	3	11	
Mesoamerica	Mexico, Central America	7.1	0	0	0	0	0	0	2	5	0	0	1	8	6	
	Races Durango/Jalisco	7.5	0	0	1	0	7	1	5	0	0	0	3	17	36	
	Race Mesoamerica	7.2	0	0	4	0	0	1	1	0	14	3	7	30	58	
	Wild	7.3	0	0	0	0	0	0	0	3	0	0	0	3	2	
	Wild	7.6	0	0	0	0	0	2	0	0	0	0	0	2	3	
		7.Other (< .80)	2	3	1	2	6	4	9	2	3	6	11	49	109	
		Total	2	6	7	2	13	8	17	10	18	10	22	115	242	26
Total analyzed in this study							357									

[a]Groupings of McClean et al. (2012) for the Andean gene pool: A:N-SA, northern South America; A:C-SA, central South America; A:S-Mx-SA, southern Mexico and South America; A:Mixed, mixed.

[b]Groupings of McClean et al. (2012) for the Mesoamerican gene pool: MA:N-Mx, northern Mexico; MA:NC-Mx, north-central Mexico; MA:SC-Mx, south-central Mexico; MA:S-Mx, southern Mexico; MA:CA, Central America; MA:N-SA, northern South America; MA:Mixed, mixed.

the two studies. For example, the McClean et al. (2012) groups MA:N-Mx ($n = 7$) and MA:SC-Mx ($n = 5$), from northern and south-central Mexico, respectively, corresponded mainly to group K7.5 (races Durango/Jalisco) of the current study. Groups MA:CA ($n = 14$, Central America) and MA:N-SA ($n = 3$, northern South America) corresponded to group K7.2 (race Mesoamerica) (Table 5).

4.3 | Evidence for hybridity of part of the germplasm

The STRUCTURE analysis provides an opportunity to assess potential hybridity of accessions and the group membership of the putative parents, as distinct from common ancestry.

For the $K = 3$ analysis (Table 3A), the "purest" group was K3.1 (Andean), based on the highest average posterior membership assignment value (henceforth membership value). The least pure was K3.3 (races Durango/Jalisco), with K3.2 (race Mesoamerica) in an intermediate position. The contributions from other groups were as follows. For K3.1, equally small contributions originated in groups K3.2 and K3.3, consistent with the contrasting origins of K3.1 (Andean) and K3.2 and K3.3 (Mesoamerican). K3.2 received contributions primarily from K3.3, consistent with the Mesoamerican nature of the two groups, but much less from K3.1, an Andean group. K3.3 received contributions from both K3.2, also a Mesoamerican group, but also from K3.1, in spite of the Andean nature of the latter. This observation suggests that members of the K3.3 group may be better combiners in crosses with other bean genotypes (as observed by Singh, Molina, Urrea, & Gutiérrez, 1993; Singh, Morales, Miklas, & Teran, 2000).

For the $K = 7$ analysis (Table 3B), the "purest" group was K7.2 (race Mesoamerica), followed in decreasing order by K7.5 (races Durango/Jalisco, Mexican accessions), K7.3 (Mexican wild beans with small seeds), K7.6 (Mexican wild beans with larger seeds), K7.4 (Andean beans from Ecuador and Peru), K7.1 (races Durango/Jalisco, Mesoamerican accessions), and K7.7 (Andean beans introduced into Mexico and Central America). As noted before for $K = 3$, the race Mesoamerica group shows little evidence of outcrossing with other groups. They appear to be a well-defined group with little affinity for other groups. Furthermore, breeding within this group (intraracial crosses) provide limited advance from selection (Singh et al., 1993), suggesting little intraracial diversity.

For each $K = 7$ group (Table 3B), contributions from other groups were generally predictable based on the genetic relatedness (Andean vs. Mesoamerican). K7.1 received more contributions from K7.5, the other races Durango/Jalisco group. This observation suggested that common ancestry rather than hybridization may be responsible for the similarity between K7.1 and K7.5. Additional contribution was made by K7.2 (race Mesoamerica). Consistent with the well-defined nature of the K7.2 group and the lack of further subdivision between $K = 3$ and $K = 7$, no group contributed more than 0.037 (group K7.1, race Durango/Jalisco, from Mexico). In group K7.3, the largest contribution was made by K7.6, the other group with wild beans from Mexico, providing another example where lower membership coefficients were due to common ancestry rather than hybridity. Group K7.7 made the largest contribution to K7.4, probably due to shared ancestry between these Andean groups. A similar picture was provided by K7.5, which received its largest contribution from K7.1, the other races Durango/Jalisco group, distributed in Mexico and Central America. In group K7.6, the largest contribution was not made by K7.3, the other group containing wild types, but by K7.5 (races Durango/Jalisco) distributed mainly in Mexico. This may suggest gene flow from domesticated to wild types, as suggested by Papa and Gepts (2003) and Papa et al. (2005). Finally, in group K7.7, the largest contribution was made by K7.4 (Andean), due to common ancestry. Other relatively important contributions made by domesticated Mesoamerican domesticated groups. The smallest contributions was made by the two wild groups.

4.4 | Representativity of the core collection

In general, there was a correspondence between the SSR analysis of McClean et al. (2012) and the current SNP-based analysis. The two studies identified the major separation in the Andean and Mesoamerican geographic gene pools and further subdivisions of the Mesoamerican gene pool into ecogeographic races, such as race Mesoamerica and races Durango/Jalisco, which had been identified previously (Kwak & Gepts, 2009; Singh et al., 1991). Concordance in membership of the different groups between the two studies may be lacking because different markers were used.

Group K3.1 largely represented race Nueva Granada of the Andean gene pool, based on the large, elongated seeds observed for many members of this group. Group 7.4 potentially represented race Peru. However, this race is rarely found outside the Andean center of origin because of photoperiod miss-adaptation, due to a poorly understood between temperature and daylength. This group may, therefore, also be representing race Nueva Granada. No representatives of race Chile (Singh et al., 1991) could be identified here. Group K3.3 represented race Durango and perhaps also race Jalisco, although negative interaction between temperature and daylength may have limited the frequency of the latter in the gene bank, outside the center of domestication (Pullman, WA: latitude of ~47° N; daylength between ~8.5 and 16 h). Seed multiplication of photoperiod-sensitive accessions would be possible only in the winter under greenhouse

conditions, assuming no interactions with temperature. Absence or at least low frequency of representatives of races Jalisco and Peru, outside their respective centers of origin in Mesoamerica and the Andes, has been observed previously (Zeven, 1997; Zeven et al., 1999).

In contrast, the presence of wild beans from Mexico seems incongruous in a core collection devoted to domesticated types. Either these wild bean accessions are removed and included in a wild *P. vulgaris* core collection or additional wild beans representing the populations distributed in Central America and the Andes are included in this core collection. Recent studies have studied diversity of wild beans (Ariani et al., 2016, 2018; Kwak & Gepts, 2009; Tohme et al., 1996).

We recommend to the National Plant Germplasm System in general, and the Western Regional Plant Introduction Station specifically, that this core collection be revised to broaden its coverage. Several approaches can be attempted to develop a more representative collection (Corak, Ellison, Simon, Spooner, & Dawson, 2019). First, separate core collections should be established for wild and domesticated types, for several reasons. Plant breeders are more likely to broaden the genetic diversity of their respective programs with domesticated types than with wild types. Wild types generally require different cultivation conditions, such as long nights to a complete reproductive cycle yielding seeds in common bean. Second, additional domesticated materials should be considered, particularly of races Jalisco and Guatemala in the Mesoamerican gene pool and races Peru and Chile in the Andean gene pool. An increase in the number of SNPs would also be desirable to facilitate genome-wide association approaches, in addition to genetic diversity and relationship studies.

ACKNOWLEDGMENTS

This research was funded by the BeanCAP project of the USDA National Institute of Food and Agriculture (NIFA) Agriculture and Food Research Initiative (AFRI) 2009-01929 to P. McClean (North Dakota State University), with subcontract to P. Gepts (University of California, Davis). We thank the Western Regional Plant Introduction Station of the USDA (Pullman, WA) for providing the core collection of *Phaseolus vulgaris*.

DATA AVAILABILITY

The SNP dataset is available for download from the University of California, Davis, Dash database: https://doi.org/10.25338/B8KP45 (Gepts et al., 2019).

AUTHOR CONTRIBUTIONS

P. Gepts was responsible for conceptualization and funding acquisition. S. Kuzay and P. Hamilton-Conaty were responsible for data acquisition. Data analyses were performed by S. Kuzay, P. Hamilton-Conaty, and P. Gepts. The original draft of the manuscript was written by S. Kuzay, and the manuscript was reviewed and edited by S. Kuzay and P. Gepts.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Paige Hamilton-Conaty https://orcid.org/0000-0001-6828-1708
Paul Gepts https://orcid.org/0000-0002-1056-4665

REFERENCES

Ariani, A., Berny Mier y Teran, J. C., & Gepts, P. (2016). Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). *Molecular Breeding*, *36*. https://doi.org/10.1007/s11032-016-0512-9

Ariani, A., Berny Mier y Teran, J. C., & Gepts, P. (2018). Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Molecular Biology and Evolution*, *35*, 119–131. https://doi.org/10.1093/molbev/msx273

Becerra Velasquez, V. L., & Gepts, P. (1994). RFLP diversity in common bean (*Phaseolus vulgaris* L.). *Genome*, *37*, 256–263. https://doi.org/10.1139/g94-036

Beebe, S., Skroch, P. W., Tohme, J., Duque, M. C., Pedraza, F., & Nienhuis, J. (2000). Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Science*, *40*, 264–273. https://doi.org/10.2135/cropsci2000.401264x

Beebe, S., Toro, O., González, A., Chacón, M., & Debouck, D. (1997). Wild-weed-crop complexes of common bean (*Phaseolus vulgaris* L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. *Genetic Resources and Crop Evolution*, *44*, 73–91. https://doi.org/10.1023/A:1008621632680

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., … Papa, R. (2013). Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist*, *197*, 300–313. https://doi.org/10.1111/j.1469-8137.2012.04377.x

Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Zeuli, P. S., … Papa, R. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, E788–E796. https://doi.org/10.1073/pnas.1108973109

Blair, M. W., Soler, A., & Cortés, A. J. (2012). Diversification and population structure in common beans (*Phaseolus vulgaris* L.). *PLoS One*, *7*. https://doi.org/10.1371/journal.pone.0049488

Bressani, R. (1993). Grain quality of common beans. *Food Reviews International*, *9*, 237–297. https://doi.org/10.1080/87559129309540960

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

Brown, A. H. D. (1989). Core collections: A practical approach to genetic resources management. *Genome*, *31*, 818–824. https://doi.org/10.1139/g89-144

Brown, A. H. D., & Spillane, C. (1999). Implementing core collections-principles, procedures, progress, problems and promise. In R. C.

Johnson & T. Hodgkin (Eds.), *Core collections for today and tomorrow* (pp. 1–9). Rome: International Plant Genetic Resources Institute.

Corak, K. E., Ellison, S. L., Simon, P. W., Spooner, D. M., & Dawson, J. C. (2019). Comparison of representative and custom methods of generating core subsets of a carrot germplasm collection. *Crop Science*, *59*, 1107–1121. https://doi.org/10.2135/cropsci2018.09.0602

Cortés, A. J., Chavarro, M., Madrinan, S., This, D., & Blair, M. W. (2012). Molecular ecology and selection in the drought-related Asr gene polymorphisms in wild and cultivated common bean (*Phaseolus vulgaris* L.). *BMC Genetics*, *13*. https://doi.org/10.1186/1471-2156-13-58

Cortés, A. J., This, D., Chavarro, C., Madrinan, S., & Blair, M. W. (2012). Nucleotide diversity patterns at the drought-related DREB2 encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *125*, 1069–1085. https://doi.org/10.1007/s00122-012-1896-5

Debouck, D. G., Toro, O., Paredes, O. M., Johnson, W. C., & Gepts, P. (1993). Genetic diversity and ecological distribution of Phaseolus vulgaris in northwestern South America. *Economic Botany*, *47*, 408–423. https://doi.org/10.1007/BF02907356

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

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

Ehrmann, J., & Ritz, K. (2014). Plant: Soil interactions in temperate multi-cropping production systems. *Plant and Soil*, *376*, 1–29. https://doi.org/10.1007/s11104-013-1921-8

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

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*, 1567–1587.

Freyre, R., Ríos, R., Guzmán, L., Debouck, D., & Gepts, P. (1996). Ecogeographic distribution of *Phaseolus* spp. (Fabaceae) in Bolivia. *Economic Botany*, *50*, 195–215. https://doi.org/10.1007/BF02861451

Gepts, P. (2000). A phylogenetic and genomic analysis of crop germplasm: A necessary condition for its rational conservation and utilization. In J. Gustafson (Ed.), *Genomes* (pp. 163–181). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-4235-3_13

Gepts, P. (2006). Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. *Crop Science*, *46*, 2278–2292. https://doi.org/10.2135/cropsci2006.03.0169gas

Gepts, P. (2014a). Domestication of plants. In R. R. B. Leakey (Ed.), *Encyclopedia of agriculture and food systems* (pp. 474–486). San Diego, CA: Elsevier. https://doi.org/10.1016/B978-0-444-52512-3.00231-X

Gepts, P. (2014b). The contribution of genetic and genomic approaches to plant domestication studies. *Current Opinion in Plant Biology*, *18*, 51–59. https://doi.org/10.1016/j.pbi.2014.02.001

Gepts, P., & Bliss, F. A. (1985). F1 hybrid weakness in the common bean: Differential geographic origin suggests two gene pools in cultivated bean germplasm. *Journal of Heredity*, *76*, 447–450. https://doi.org/10.1093/oxfordjournals.jhered.a110142

Gepts, P., & Bliss, F. A. (1986). Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. *Economic Botany*, *40*, 469–478. https://doi.org/10.1007/BF02859660

Gepts, P., Kuzay, S., & Hamilton, P. (2019). An evaluation of the representativity of the USDA core collection of common bean as assessed by SNP diversity, UC Davis, dataset. *Dryad*, https://doi.org/10.25338/B8KP45

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

Hardarson, G., Bliss, F., Cigales Rivero, M., Henson, R., Kipe-Nolt, J., Longeri, L., … Tsai, S. M. (1993). Genotypic variation in biological nitrogen fixation in common bean. *Plant and Soil*, *152*, 59–70. https://doi.org/10.1007/BF00016333

Hodgkin, T., Brown, A. H. D., van Hintum, T. J. L., & Morales, E. A. V. (1995). *Core collections of plant genetic resources*. Rome: International Plant Genetic Resources Institute.

Hufford, M. B., Berny Mier y Teran, J. C., & Gepts, P. (2019). Crop biodiversity: An unfinished magnum opus of nature. *Annual Review of Plant Biology*, *70*, 727–751. https://doi.org/10.1146/annurev-arplant-042817-040240

Hyten, D. L., Song, Q. J., Fickus, E. W., Quigley, C. V., Lim, J. S., Choi, I.-Y., … Cregan, P. B. (2010). High-throughput SNP discovery and assay development in common bean. *BMC Genomics*, *11*, https://doi.org/10.1186/1471-2164-11-475

Kami, J., Becerra Velásquez, B., Debouck, D. G., & Gepts, P. (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proceedings of the National Academy of Sciences of the United States of America*, *92*, 1101–1104. https://doi.org/10.1073/pnas.92.4.1101

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

Koenig, R., Singh, S. P., & Gepts, P. (1990). Novel phaseolin types in wild and cultivated common bean (*Phaseolus vulgaris*, Fabaceae). *Economic Botany*, *44*, 50–60. https://doi.org/10.1007/BF02861066

Kwak, M., & Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics*, *118*, 979–992. https://doi.org/10.1007/s00122-008-0955-4

Kwak, M., Kami, J. A., & Gepts, P. (2009). The putative Mesoamerican domestication center of Phaseolus vulgaris is located in the Lerma-Santiago basin of Mexico. *Crop Science*, *49*, 554–563. https://doi.org/10.2135/cropsci2008.07.0421

Kwak, M., Toro, O., Debouck, D., & Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris* L.). *Annals of Botany*, *110*, 1573–1580. https://doi.org/10.1093/aob/mcs207

Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., & McClean, P. E. (2013). Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity*, *110*, 267–276. https://doi.org/10.1038/hdy.2012.82

McClean, P. E., Terpstra, J., McConnell, M., White, C., Lee, R., & Mamidi, S. (2012). Population structure and genetic differentiation

among the USDA common bean (*Phaseolus vulgaris* L.) core collection. *Genetic Resources and Crop Evolution*, *59*, 499–515. https://doi.org/10.1007/s10722-011-9699-0

Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., … Papa, R. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSHP1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, *123*, 1341–1357. https://doi.org/10.1007/s00122-011-1671-z

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

Papa, R., Acosta, J., Delgado-Salinas, A., & Gepts, P. (2005). A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theoretical and Applied Genetics*, *111*, 1147–1158. https://doi.org/10.1007/s00122-005-0045-9

Papa, R., Bellucci, E., Rossi, M., Leonardi, S., Rau, D., Gepts, P., … Attene, G. (2007). Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany*, *100*, 1039–1051. https://doi.org/10.1093/aob/mcm151

Papa, R., & Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics*, *106*, 239–250. https://doi.org/10.1007/s00122-002-1085-z

Paredes, M., Becerra Velásquez, V. L., Tay Urbina, J., Blair, M. W., & Bascur, G. (2010). Selection of a representative core collection from the Chilean common bean germplasm. *Chilean Journal of Agricultural Research*, *71*, 3–13. https://doi.org/10.4067/S0718-58392010000100001

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research: An update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

Rendón-Anaya, M., Herrera-Estrella, A., Gepts, P., & Delgado-Salinas, A. (2017). A new species of Phaseolus (Leguminosae, Papilionoideae) sister to *Phaseolus vulgaris*, the common bean. *Phytotaxa*, *313*, 259–266. https://doi.org/10.11646/phytotaxa.313.3.3

Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Alvarez, S., Vlasova, A., Capella-Gutiérrez, S., Ordaz-Ortiz, J. J., … Herrera-Estrella, A. (2017). Genomic history of the origin and domestications of common bean in the Americas unveils its closest sister species. *Genome Biology*, *18*, https://doi.org/10.1186/s13059-017-1190-6

Schmutz, J., McClean, P., Mamidi, S., Wu, G., Cannon, S., Grimwood, J., … Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*, 707–713. https://doi.org/10.1038/ng.3008

Singh, S. P., Gepts, P., & Debouck, D. G. (1991). Races of common bean (*Phaseolus vulgaris* L., Fabaceae). *Economic Botany*, *45*, 379–396. https://doi.org/10.1007/BF02887079

Singh, S. P., Molina, A., Urrea, C. A., & Gutiérrez, J. A. (1993). Use of interracial hybridization in breeding the race Durango common bean. *Canadian Journal of Plant Science*, *73*, 785–793. https://doi.org/10.4141/cjps93-101

Singh, S. P., Morales, F. J., Miklas, P. N., & Teran, H. (2000). Selection for bean golden mosaic resistance in intra- and interracial bean populations. *Crop Science*, *40*, 1565–1572. https://doi.org/10.2135/cropsci2000.4061565x

Singh, S. P., Nodari, R., & Gepts, P. (1991). Genetic diversity in cultivated common bean. I. Allozymes. *Crop Science*, *31*, 19–23. https://doi.org/10.2135/cropsci1991.0011183X003100010004x

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

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*, 512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–2739. https://doi.org/10.1093/molbev/msr121

Tohme, J., González, D. O., Beebe, S., & Duque, M. C. (1996). AFLP analysis of gene pools of a wild bean core collection. *Crop Science*, *36*, 1375–1384. https://doi.org/10.2135/cropsci1996.0011183X003600050048x

Upadhyaya, H. D., Wang, Y.-H., Gowda, C. L. L., & Sharma, S. (2013). Association mapping of maturity and plant height using SNP markers with the sorghum mini core collection. *Theoretical and Applied Genetics*, *126*, 2003–2015. https://doi.org/10.1007/s00122-013-2113-x [erratum: 127(6):1461].

Zeven, A. C. (1997). The introduction of the common bean (Phaseolus vulgaris L.) into Western Europe and the phenotypic variation of dry beans collected in the Netherlands in 1946. *Euphytica*, *94*, 319–328. https://doi.org/10.1023/A:1002940220241

Zeven, A. C., Waninge, J., van Hintum, T., & Singh, S. P. (1999). Phenotypic variation in a core collection of common bean (*Phaseolus vulgaris* L.) in the Netherlands. *Euphytica*, *109*, 93–106. https://doi.org/10.1023/A:1003665408567

SUPPORTING INFORMATION

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

How to cite this article: Kuzay S, Hamilton-Conaty P, Palkovic A, Gepts P. Is the USDA core collection of common bean representative of genetic diversity of the species, as assessed by SNP diversity? *Crop Science*. 2020;60:1398–1414. https://doi.org/10.1002/csc2.20032