THEORETICAL AND APPLIED GENETICS
International Journal of Plant Breeding Research
Springer

Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.).

Journal:	*Theoretical and Applied Genetics*
Manuscript ID:	draft
Manuscript Type:	Original Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Blair, Matthew; CIAT, Biotechnology Unit Díaz, Juan; CIAT, Biotechnology Hidalgo, Rigoberto; Univ. Nacional, Agronomy Díaz, Lucy; CIAT, Biotechnology Unit Duque, Myriam; CIAT, Biotechnology Unit
Keywords:	common bean races, population structure, gene flow, simple sequence repeat markers

scholarONE™
Manuscript Central

For Review

Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.).

M. W. Blair[1]✉, J.M. Díaz[1,2], R. Hidalgo[2], L. M. Díaz[1], M.C. Duque[1]

(1) Centro Internacional de Agricultura Tropical (CIAT), Apartado Aéreo 6713, Cali, Colombia

(2) Universidad Nacional de Colombia, Palmira, Colombia

Received:

Corresponding author: ✉ M. W. Blair

CIAT - International Center for Tropical Agriculture
1380 N.W. 78th Ave., Miami, Florida 33126 USA or
A. A. 6713, Cali, Colombia, South America
Tel: 57-2-4450-000; VIA USA 650-833-6625
Fax: 57-2-4450-073; VIA USA 650-833-6626 or 305-592-9757

E-mail: m.blair@cgiar.org

For Peer Review

Abstract

The Andean gene pool of common bean (*Phaseolus vulgaris* L.) has high levels of morphological diversity in terms seed color and size, growth habit and agro-ecological adaptation, but previously was characterized by low levels of molecular marker diversity. Three races have been described within the Andean gene pool: Chile, Nueva Granada and Peru. The objective of this study was to characterize a collection of 123 genotypes representing Andean bean diversity with 33 microsatellite markers that have been useful for characterizing race structure in common beans. The genotypes were from both the primary center of origin as well as secondary centers of diversity to which Andean beans spread and represented all three races of the gene pool. In addition we evaluated a collection of landraces from Colombia to determine if the Nueva Granada and Peru races could be distinguished in genotypes from the northern range of the primary center. Multiple correspondence analyses of the Andean race representatives identified two predominant groups corresponding to the Nueva Granada and Peru races. Some of the Chile race representatives formed a separate group but several that had been defined previously as from this race grouped with the other races. Gene flow was more notable between Nueva Granada and Peru races than between these races and the Chile race. Among the Colombian genotype, the Nueva Granada and Peru races were identified and introgression between these two races was especially notable. The genetic diversity within the Colombian genotypes was high, reaffirming the importance of this region as an important source of germplasm. Results of this study suggest that the morphological classification of all climbing beans as Peru race genotypes and all bush beans as Nueva Granada race genotypes is erroneous and that growth habit traits have been mixed in both races, requiring a re-adjustment in the concept of morphological races in Andean beans.

For Peer Review

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the most important and diverse legume staples consumed worldwide (Broughton et al. 2003). The genetic resources of common bean are sub classified into races that are made up of morphologically-similar cultivars that share equivalent agro-ecological adaptation and some agronomic characteristics such as growth habit and seed type (Singh et al. 1991a). Seven races of cultivated common bean have been described and are grouped according to which of two gene pools they belong to, with the Andean gene pool originating in the highlands of South America and the Mesoamerican gene pool originating in Central America and Mexico (Gepts 1988; Singh 1989). Andean beans can be distinguished from Mesoamerican beans on the basis of morphological and biochemical differences such as seed size and seed proteins as well as with various types of molecular markers (Gepts 1998). Meanwhile the races within each gene pool can be distinguished somewhat by some but not all molecular markers (Beebe et al. 2000, 2001; Diaz and Blair 2006) and to a certain extent by allelic frequency at a limited number of isozyme loci (Koenig and Gepts 1989; Singh et al. 1991c). Germplasm exchange, gene flow and crossing programs between and within the two gene pools have given rise to introgression between gene pools (Beebe et al. 2001; Islam et al. 2004) and between races (Diaz and Blair 2006) resulting in intermediates phenotypes that do not classify well into any single race or gene pool. Furthermore, it is not known whether races originated from separate multiple domestications or were the result of a single domestication in each region with subsequent differentiation and adaptation to specific agroecological niches (Chacón et al. 2005).

While DNA polymorphisms between Andean and Mesoamerican gene pools have been well characterized (Becerra and Gepts 1994, Gepts 1998) differences between the races within each of the gene pools have been less so. This has been especially true for Andean beans where previous marker technologies were unsuccessful at detecting race structure. This was the case for Johns et al. (1997) who tried to distinguish the Chile race using RAPDs, Duarte et al. (1999) who attempted to differentiate various races with RAPDs and

For Peer Review

Beebe et al. (2001) who used AFLPs to characterize wild and cultivated Andean beans but could only distinguish wild accessions not cultivars. The conclusions of these three studies predict that there is less diversity among Andean cultivars than in Mesoamerican cultivars and that compared to wild accessions the cultivated Andean gene pool underwent a reduction in diversity, however none of the studies could define race structure at the molecular level or define how this structure would agree with the morphological classification made by Singh et al. (1991a, b). Recent results from our laboratory for a sample of breeding lines and germplasm accessions show that microsatellite markers uncover high levels of diversity in Andean beans and that this diversity may be higher even than in the Mesoamerican gene pool (Blair et al. 2006). These results agree with a study in snap beans, which are predominantly Andean, where microsatellites were used to explore diversity among different commercial classes (Metais et al. 2002). In both studies microsatellite markers were advantageous because they detected greater genetic polymorphisms compared to other marker types. Microsatellites have also been useful in the definition of race structure in the Mesoamerican gene pool (Diaz and Blair 2006) and this study is an extension of that work in the analysis of the Andean race structure.

The specific objective of this study, therefore, was to determine race structure in Andean beans based on a sample of 123 landraces representing the three morphologically defined races of Andean beans (Nueva Granada, Peru and Chile) analyzed with microsatellite markers that have been useful for characterizing overall diversity and Mesoamerican race structure in our previous studies (Blair et al. 2006; Diaz and Blair 2006). The genotypes were selected from both the primary center of origin including Bolivia, Peru and Ecuador as well as secondary centers of diversity to which Andean beans spread. A collection of landraces from Colombia was used to determine if the Nueva Granada and Peru races could be distinguished in genotypes from the northern range of the primary center. In addition to determining race structure *per se* we were also interested in determining the correlation of race structure with phenotypic and morphological differences such as growth habit and seed characteristics.

For Peer Review

Materials and Methods:

Plant Material: A total of 123 genotypes were evaluated in this study and were selected based on previous race designation. The study was conducted in two phases with two sets of genotypes (Table 1). The first group consisted in 63 genotypes and included representatives of races Nueva Granada, Peru and Chile selected from accessions previously classified as belonging to one or the other of these races by Singh et al. (1991b, c) or Becerra and Gepts (1994). Within this group 28 accessions were selected to be from race Peru, 25 from race Nueva Granada and 6 from race Chile along with 4 control genotypes described below. Race Peru genotypes came from Argentina (1), Colombia (7), Ecuador (8) and Peru (12); while race Nueva Granada genotypes came from Argentina (1), Brazil (4), Colombia (13), Cuba (1), Ecuador (2), Mexico (2), Dominican Republic (1) and Yugoslavia (1). All race Chile genotypes were from Chile while race Peru genotypes included 8 popping beans (Ñuñas). The majority of the genotypes in this group were landraces although a few were improved varieties (4). Most but not all of the genotypes were selected from the core collection developed at CIAT (Tohme et al. 1995a) and were evaluated previously by Beebe et al. (2001). The second group consisted of 60 accessions from Colombia (50 native landraces, 2 commercial varieties and 8 advanced lines from CIAT) all of which are found in the CIAT core collection and have Andean phaseolin patterns along with the 4 control genotypes described below. The Colombian landraces represented all the major producing regions of the country with 14 accessions from Antioquia, 7 from Boyacá-Cundinamarca, 11 from Cauca-Huila-Tolima, 4 from Caldas-Valle, 16 from Nariño and 8 from the CIAT and CIAT/ICA collaborative breeding programs. A total of 4 genotypes were used as controls across the two groups since they had been used in our previous microsatellite studies (Blair et al. 2006; Diaz and Blair 2006): 'Calima' from Colombia (G4494), 'Chauca Chuga' from Peru (G19833) as Andean control genotypes and DOR364 also known as the variety 'Dorado' from CIAT / El Salvador and ICA Pijao from Colombia used as Mesoamerican outgroup individuals. All seeds were provided by the Genetic Resources Unit of CIAT.

For Peer Review

DNA extraction and phaseolin evaluation: DNA extraction involved germinating 10 seeds selected at random from each accession, which were scarified to ensure uniform germination in the laboratory and pre-germinated in darkness on germination paper. The first trifoliate leaves of eight-day old seedlings were collected and ground in liquid nitrogen for DNA extraction with the method of Afanador et al. (1993). DNA was resuspended in TE buffer and DNA quality was evaluated on 0.8% agarose gels followed by quantification with Hoescht H 33258 dye on a Hoefer DyNA fluorometer (DNA Quant™ 200). DNA was diluted to 10 ng/µl for further experiments. The phaseolin pattern of each genotype was either known from previous studies (Singh et al. 1991c) or evaluated by SDS-PAGE electrophoresis of total seed proteins as described in Durán et al. (2005).

Microsatellite amplifications: In all, 33 microsatellites were used of which 11 were cDNA based and 22 were genomic. Microsatellites were selected based on their distribution in the genome (average of three per linkage group) and on their polymorphism information content (Blair et al. 2003; 2006). The genomic microsatellites included BM139, 140, 142, 143, 152, 155, 160, 164, 167, 170, 175, 183, 184, 189, 197, 205; BMd12, 33; and GATs11B, 54, 91. The genic microsatellites included BMc5; BMd10, 15, 17, 20, 26, 28; and PV-ctt001, -ag001, -ag003, -gaat001 (Blair et al. 2003; Gaitán et al. 2002; Yu et al. 2000). PCR amplifications were conducted in 96-well plates using a PTC 100(MJ Research) thermal cycler with conditions as given in Blair et al. (2006). PCR reaction volume was 12 µl and contained 50 ng genomic DNA and 0.16 µM of each primer (forward and reverse), 10 mM Tris-HCl (pH 7.2), 50 mM KCl, 1.5 to 2.5 mM Mg (depending on the primer), 0.2 mM dNTP and 1.0 unit Taq polymerase. PCR products were elecrophoresed on 4% denaturing polyacrylamide gels using Sequi-Gen® GT sequencing units (Bio-Rad Laboratories, Inc., Hercules, CA, USA) using 0.5 X TBE buffer at a constant 100 W for approximately 1.5 hours. Gels were stained with silver nitrate according to manufacturer's instructions (Promega Inc., Madison, WI) and allele sizes were evaluated relative to a 10 bp molecular weight size standard (Invitrogen, Carlsbad, CA). Allele sizes for the check genotypes (Calima, DOR364, ICA Pijao and G19833) were confirmed to be the same sizes as in Blair et al. (2006).

For Peer Review

Data analysis: The allele information coded for band presence or absence was used for multiple correspondence analysis in NTSYS-pc v. 2.10 (Rohlf, 2002) from which a matrix of Euclidean distances between genotypes was calculated with the Corresp and Simint subprograms. The distance matrix was then used in the SAHN subprogram for cluster analysis to construct dendrograms based on UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering method which were visualized with Tree plot. In addition, the coordinates for each accession in the multiple correspondence analysis were plotted using the Graph module and the G3D procedure of the software program SAS (SAS-Institute, 1996). Within group genetic diversity was estimated for each group found within the dendrogram based on the genetic distance between all pairs of genotypes (Nei 1978) using the software program POPGENE version 1.31 (Yeh et al. 1997). Other common parameters of genetic diversity (observed heterozygosity (Ho), indices of genetic differentiation (Gst), and gene flow (Nm) were determined with the same software. Genetic differentiation was estimated from the formula from Nei (1987): $G_{ST}=D_{ST}/H_T$; where $H_T=H_S+D_{ST}$ and $D_{ST}=H_T-H_S$ based on H_T being the total genetic diversity for the populations, H_S is the within subpopulation genetic diversity and D_{ST} is the between subpopulation genetic diversity. Gene flow (Nm) was equivalent to the number of migrants between populations per generation calculated based on the formula $Nm=0.25(1-G_{ST})/G_{ST}$ (Slatkin and Barton 1989). A prevalence index was estimated according to the formula: $PI = \sum M_i * F_i / \sum M_i$ where M_i = marker genotype for each individual for the i^{th} allele coded for presence (1) or absence (0) and F_i = frequency of i^{th} allele in the population (Beer et al. 1997). The number of populations (K) was confirmed with the software STRUCTURE (Pritchard 2000) and visualized with the software DISTRUCT (Rosenberg et al. 2002). Population structure was determined assuming an admixture model with K=3 and K=4 based on prior population information from the multiple correspondence analysis and a total of 30,000 iterations each for both MCMC repetitions and burn in.

For Peer Review

Results

Multiple correspondence analysis with Andean race representatives

Four groups could be distinguished in the multiple correspondence analysis performed with the global sample of Andean race representatives (Figure 1). The largest groups corresponded to the Nueva Granada and Peru races with 30 and 25 genotypes each, respectively. These two groups included the two genotypes used as Andean controls, namely Calima (G4494) found in the Nueva Granada group, and Chauca Chuga (G19833), found in the Peru race group. Another group corresponded to three of the Chile race representatives (G4472, G4474 and G5831) along with one (G13940) of the race Peru genotypes. Meanwhile, a final group corresponded to the Mesoamerican control genotypes (DOR364 and ICA Pijao) along with two Brazilian genotypes classified by Singh et al. (1991b) as race Nueva Granada, both of the Jalo commercial class (medium seeded varieties with T phasoeolin, G5204 and G21953). These are likely to show introgression with Mesoamerican genotypes given their intermediate position between the two gene pools. The four groups were separated on different axis with Andean versus Mesoamerican control genotypes (and Jalo accessions) separated in the first dimension, with Nueva Granada and Peru race genotypes separated in the second dimension and with race Peru and race Chile separated on the third dimension.

The results show that all the genotypes classified by Singh et al. (1991b, c) as race Nueva Granada continued in this group while several genotypes classified by these authors as race Peru genotypes actually belonged to race Nueva Granada in our study. These included the genotypes G111 (Huasca Colorado), G5702 (Cargamanto), G5708 (Sangretoro), G6379 (ICA Tone) and G7231 (Antioquia 47/Cargamanto), mostly climbing beans but with grain color and shape similar to other Nueva Granada race genotypes. Similarly three genotypes from the Frutilla and Coscorron commercial classes classified by Singh et al. (1991b, c) as belonging to race Chile actually belonged to race Nueva Granada (genotypes G18356 and G18372) or to race Peru (G2544) in this study. Furthermore, several genotypes previously

For Peer Review

classified as race Peru genotypes by Singh et al. (1991b, c) were intermediate between the different clusters, including G23604 (Nuña/popping bean) which was located between the race Peru and race Chile groups; and G12644 (Morado) which was located between the race Peru and race Nueva Granada groups. Meanwhile the genotype G7930 (Alubia) while in the Nueva Granada group showed some signs of introgression with the Mesoamerican group in agreement with previous reports by Islam et al. (2004).

Genotype associations in the Andean gene pool

The dendrogram produced for the Andean race representatives (Figure 2), shows the same results as the correspondence analysis, with clear separation of races Chile, Nueva Granada and Peru at a Euclidean distance of 0.85 along with separation of the Mesoamerican control genotypes clustered with the two Jalo accessions as described above at a Euclidean distance of 2.80. The Andean control genotypes, G4494 (Calima) and G19833 were found in the clusters representing the Nueva Granada and Peru races, respectively. In addition to showing these divisions between the races, the dendrogram also shows the close similarity of related genotypes from either the primary or secondary centers of diversity for Andean beans. For example within the group that makes up the Nueva Granada race, two yellow-seeded Mexican genotypes (G1326 and G2488) appeared on the same branch. Both of these genotypes are of the "Canario" commercial class. Similarly the Brazilian genotypes Bagajó (G5254) was on the same branch as Jalo EEP558 (G9603). Two varieties from the Caribbean, Velasco Largo from Cuba and Pompadour Checa 50 from the Dominican Republic (G18255 and G18264, respectively) were also closely related. The Nueva Granada genotypes from secondary centers of diversity clustered with other kidney shaped beans from the primary center of diversity in the Andean region, including Colombian red and red mottled varieties such as Calima (G4494), ICA Toné (G6379), ICA Guali (G4452) and Sangretoro (G5708). It was notable that all the yellow seeded beans, both from secondary centers of diversity as well as from the Andean region were in this overall cluster showing an association with the red and red mottled kidney shaped beans.

For Peer Review

Another cluster that was distinct from the previous set of accessions with kidney shaped seeds formed around the red seeded "Radical" genotypes (G4080 and G4717) that grouped with other round-seeded Colombian beans (G14660 and G4520). Similarly, the Cargamanto genotypes, previously thought to be from race Peru clustered with the Nueva Granada genotypes Cargabello and Uribe Redondo (G21272 and GG8147, respectively), all with large, rounded, cream to pink mottled seed types. These results suggest the formation of two subraces within the Nueva Granada group separating at Euclidean distance of 0.29 (clusters A and B in the dendrogram). Also within the Nueva Granada race but apart from these previous two subgroups at a Euclidean distance of 0.60, the introgressed genotype G7930 (Alubia) clustered with two genotypes also showing large elongated seed, G7309 (Sangretoro) and G21227 (Mortiño Alargado), all of which have "T" phaseolin (cluster C in the dendrogram).

Within the Peru race group there was also evidence of two subraces, separating at Euclidean distance of 0.58. The first subgroup (cluster A in the dendrogram) consisted in a mixture of type III and type IV growth habit genotypes with somewhat elongated to round shaped grain from various commercial classes including "Caballero" (G5723 and G18325) and "Liborino" (G11819 and G19833) beans, in addition to the Chilean genotype G2544. The second subgroup (cluster B) consisted in type IV growth habit climbing beans with large round seeds of various seed colors including those identified as "Bola Roja" or "Bolon Rojo" (G12209, G12229, G12421), "Bolon Bayo" (G12407), "Canario" (G12207), and "Mortiño" (G12709). The popping beans or ñuñas were distributed among both subgroups rather than being concentrated in either subgroup exclusively.

Phaseolin pattern, seed size, growth habit and race designation

The agreement between race designation and phaseolin pattern is shown in Table 2 while the correlation of seed size and growth habit with each of the races is shown in Table 3. The Nueva Granada and Peru genotypes had more phaseolin alleles than race Chile. Although race Peru genotypes had wide variability in phaseolin alleles (C, H, Pa/T, T,

For Peer Review

T/H1), the majority of genotypes had either "H" or "T" phaseolin. Meanwhile, most race Nueva Granada genotypes had the "T" phaseolin allele expected for the race, but other alleles (C, Ca, H and T/H1) not considered typical of the race by Singh et al. (1991a) were also found. Among the Nueva Granada race genotypes wide variability of phaseolin alleles was found for genotypes from Colombia. Race Chile genotypes meanwhile had a mixture of phaseolin alleles (Pa, T and To) but not those thought to be characteristic of the race by Singh et al. (1991a), such as phaseolin "C" and "H".

In terms of seed size, race Nueva Granada genotypes had either medium (37.9%) or large (62.1%) seed size, while race Peru had a mixture of small (4.2%), medium (16.7%) and large (79.2%) seed size. It was notable that within the Peru race the medium and small seed size was present for round seeded popping and non-popping beans such as "Caballero" and "Huevo de Pinche" types while within the Nueva Granada race the medium seed size was present for "Canario", "Jalo", "Cargabello" and "Calabozo" varieties. Growth habit varied within both the Nueva Granada and Peru races, with the first race showing mostly type I determinates (79.3%) but also type III and type IV indeterminates (20.7%); and the second race showing mostly type IV climbing beans (81.8%) but also some type I and type III bush beans (18.2%).

Race structure analysis with Colombian genotypes

The dendrogram for the Colombian genotypes (Figure 2) shows two major groups at a Euclidean distance of 1.1 corresponding to the Nueva Granada race with 41 genotypes and the Peru race with 16 genotypes. Each of these groups was associated with one of the control genotypes discussed above; with G4494 (Calima) found within the Nueva Granada race and G19833 found within the Peru race. Within the Nueva Granada race there were three groups that separated at a Euclidean distance of 0.72. The first group of 18 genotypes (cluster A in the dendrogram) contained mostly type I growth habit bush beans (64.7%) along with two each of type II bush beans, type III semi-climbing beans and type IV climbing beans (total of 35.3%) and was made up of a range of improved breeding lines

For Peer Review

(ABA2, AND1005, BAT1373, LRK31, PVA773 and PVA1111), landraces (G4564, G4644, G4653, G4672, G4681, G4780, G4906, G21210, G21227 and G23682) along with the control genotype Calima (G4494). The second group (cluster B) also consisted of 18 genotypes however in this case all were landrace varieties and most was type IV climbing beans (78%) with only a few type I bush beans (22%). The third group (cluster C) consisted in five genotypes, four of which were type III or type IV growth habit and only one of which was type I growth habit. This group separated from the previous two groups at a Euclidean distance of 0.85 and was intermediate in morphology between the Nueva Granada and Peru race genotypes. Meanwhile the Peru race included mostly type IV climbing beans (62.5%) but also two type III growth habit genotypes (the control genotypes Chauca Chuga and G14016) and four genotypes (25.1%) with either type II or type I growth habits including the breeding lines CAL149 and DRK47 and the landraces G4547 and G13220. Six additional genotypes (G4691, G7314, *G7257*, *G8157*, G12667 and *G19142B*) fell outside of the groups for race Peru or race Nueva Granada. This group represented introgression from the Mesoamerican gene pool as indicated by their intermediate position between the Mesoamerican control genotypes and the two Andean races and given that three of the genotypes (those in italics in the list above) were classified as introgressed in the study of Islam et al. (2004).

Genetic diversity and population structure

Among the Andean race representatives we found a total of 166 alleles for the markers evaluated (average of 5.7 alleles per locus) while for the Colombian genotypes we found a total of 216 alleles for the markers evaluated (average of 7.2 alleles per locus). The most polymorphic markers were BM143 (14 and 23 alleles, respectively), BM152 (12 and 13 alleles), BM160 (13 and 18) and BM167 (18 and 12). It was notable that the cDNA-based microsatellites were less polymorphic (averages of 3.7 and 4.4 alleles) per locus than the genomic microsatellites (averages of 6.9 and 8.6 alleles). The genetic diversity based on Nei's index for the entire set of Andean race representatives was 0.49 while for the Colombian genotypes it was 0.54. Among the Andean race representatives, genetic diversity was highest for the Peru (0.39) and Nueva Granada (0.38) races and lower for

For Peer Review

race Chile (0.32); and genetic differentiation (G_{ST}) between the Nueva Granada and Peru races was low (0.093) while it was higher for race Chile compared to both race Nueva Granada (0.144) and race Peru (0.170). Correspondingly there was high gene flow (Nm) between race Nueva Granada and race Peru (2.45) but lower gene flow between these two races and race Chile (1.49 and 1.22, respectively). Among the groups identified for the Colombian genotypes, genetic differentiation (G_{ST}) was low averaging 0.10 between race Peru and the three subgroups of race Nueva Granada; while gene flow (Nm) was high and averaged 2.38 between these same groups. Observed heterozygosity across all microsatellites was low being 0.065 for the Andean race representatives and 0.027 for the Colombian genotypes. Graphs of the prevalence indices (Figure 3) for the Andean race representatives showed that among the three races, race Chile genotypes had the less prevalent alleles on average (0.48), while race Nueva Granada and race Peru genotypes had similar average allele prevalence (0.55 and 0.52, respectively). The average prevalence index was lower for the Mesoamerican control genotypes and introgressed accessions (0.27). For the Colombian genotypes, the same analysis showed similar average prevalence indices for the Nueva Granada and Peru race genotypes (0.47 and 0.46 respectively) showing that within Colombian germplasm many alleles are shared between the races and gene flow has occurred between them. Introgressed individuals among the Colombian genotypes had lower prevalence indices compared to the two races within this germplasm set. Population structure analysis for the Andean race representatives showed good agreement with the multiple correspondence analysis assignment of genotypes to races with clear distinction between race Peru and race Nueva Granada at K=3 (Figure 4) however, race Chile genotypes grouped with race Peru genotypes rather than being separate. At K=4, race Chile genotypes were associated with a group of race Nueva Granada genotypes and with admixture in some race Peru genotypes. Population structure analysis for the Colombian genotypes showed significant admixture between the races especially at K=3. At K=4, race Peru and race Nueva Granada genotypes were clearly identified and the subgroups described for Nueva Granada (NG-A, NG-B and NG-C) could be distinguished. The introgressed genotypes with similarity to the Mesoamerican controls were easily distinguished in both germplasm sets at both K=3 or K=4.

For Peer Review

Discussion

Our results use microsatellite markers to show definitively that race structure exists in the Andean gene pool confirming partially the results of Singh et al. (1991b and c) that were based on agro-morphological and isozyme based classification. The success of microsatellite markers in detecting race structure in the Andean gene pool contrasts with studies using RAPD and AFLP markers that have not been able to separate the Andean races (Duarte et al. 1999; Beebe et al. 2001; Islam et al. 2004) even though some evidence for Andean race structure was evident in RFLPs analysis (Becerra and Gepts 1994) and in our previous study with microsatellites of a limited number of Andean genotypes (Blair et al. 2006). The analysis shown here is more definitive due to the large number of genotypes analyzed and the widespread distribution across the genome of loci sampled. Although the results of correspondence analysis for the Andean race representatives clearly showed the separation of the three Andean races proposed by Singh et al. (1991a), namely Nueva Granada, Peru and Chile; there were differences in the assignment of specific genotypes from the earlier studies we based our selection of genotypes on. These discrepancies were illustrative of the limitations of using agro-morphology and origin for classification as has been pointed out before by Singh et al. (1991b). It was notable that we found several climbing bean genotypes thought previously to be of the Peru race were actually of the Nueva Granada race and vice versa several determinate bush beans previously thought to be of the Nueva Granada race could be classified as of the Peru race. Similarly, several representatives thought to be of race Chile were actually from the Nueva Granada or Peru race. Further findings from this study can be divided into specific observations about each of the three races described below and comments on introgression between these as was found for Colombian genotypes.

Race Chile, the most minor of the three races, was the most difficult to distinguish in this study and was supported by a cluster of only four individuals in the multiple correspondence analysis. These genotypes shared the characteristics of all being indeterminate type III growth habit beans with medium to large tan seed. Three of the

For Peer Review

genotypes were from Chile and belonged to the Tortola and Coscorron seed class, while one Argentinean genotype (G13490) that grouped with these genotypes was an Overito type. The grouping of both Chilean and an Argentinean landraces indicates a possible link between genotypes from the southern range of wild and cultivated common beans in the Andes and would provide a possible explanation for the origin of the race from wild beans that do exist in Argentina but that do not exist in Chile (Santalla et al. 2004). This point is important given the lack of evidence for a separate domestication leading to race Chile and the incongruence that this race may have arisen separately even though it has not been associated with specific wild accessions (Debouck et al. 1993; Johns et al. 1997; Chacón et al. 2005). Our results may indicate that race Chile precursors moved from the primary center of origin in Northwestern Argentina across the Andes mountains into temperate latitudes in central Chile, a hypothesis that would be supported by results from Sonnante et al. (1994) showing close relationship of race Chile genotypes with Argentine wild accessions. This would require additional study using wild accessions to determine if the some contribute to race Chile and are distinct from those wild beans contributing to races Nueva Granada and Peru that are thought to be in the primary center of origin especially in Bolivia and Southern Peru (Beebe et al. 2001). Our results showing similarity of race Chile genotypes with race Peru in both the multiple correspondence and population structure analyses, validate the observation that some race Chile accessions are found outside of Chile in Argentina, Southern Peru and Bolivia (Singh et al. 1991a). Race Peru isozyme alleles have been found in Chilean genotypes by Paredes and Gepts (1995) who like us found that race Chile accessions have some distinct alleles compared to other Andean genotypes. However their conclusion that Chilean genotypes might have introgression with the Mesoamerican gene pool was not supported by our study although we did find some gene flow and population admixture between race Chile and the other Andean races including both race Peru and race Nueva Granada. It is evident from the studies involving Chilean germplasm that the mixture of alleles present indicate the need to redefine the concept of race Chile, predicting that the germoplasm from this region represents an amalgam of imported germplasm with components of a unique gene pool from the southern Andes and Argentina where diverse wild beans are found (Santalla et al. 2004).

For Peer Review

Race Nueva Granada, the most important of the three races in terms of commercial importance and worldwide presence, was originally thought to be made up of bush bean genotypes (Singh et al. 1991a) but as mentioned above one notable discovery of this research was that this race also includes climbing beans as well as bush beans. Results from our study suggest that Nueva Granada race genotypes are distinguished from Peru race genotypes by producing elongated, cylindrical or slightly oval seed such as those of the Calima, Dark Red Kidney, Jalo commercial classes versus the rounded seed typically found in popping beans and other Peru race seed types. Nueva Granada race genotypes also differ in terms of their adaptation range and can be found in a wide range of tropical, subtropical and temperate regions compared to the more limited environment where race Peru genotypes are grown at high elevations sites in the Andes. The wider adaptive range in the Nueva Granada race has led to their importance in many parts of the world including in Brazil, Canada, Mexico and the United States, as well as countries of Asia, Eastern or Southern Africa and Northern or Southern Europe. It also appears that introgression with the Mesoamerican gene pool has been more common with race Nueva Granada genotypes such as with the Jalo group in the Brazilian secondary center of diversity where Mesoamerican beans are found and gene flow may be common (Maciel et al. 2003). Meanwhile, gene flow between the Nueva Granada and Peru races is likely to be common in the Andes in regions where the two races overlap as in Ecuador and Colombia and leads to intermediate genotypes such as those we observed in this study. Some race Nueva Granada genotypes have similar seed colors to those found in race Peru, notably for the red-seeded Radical and Sangretoro types. These intermediate forms may also be leading to the formation of subraces for both the Nueva Granada and Peru races although this will need to be confirmed with further studies.

Race Peru, less important than race Nueva Granada but more widely distributed than race Chile, was thought to be all type IV climbing beans by Singh et al. (1991a) but here we show the race has type III and even type I bush beans and is made up of genotypes with both patterned and un-patterned seeds that tend to be rounded or oval such as is found in

For Peer Review

the Caballeros, Bola Roja, Canario and Liborino commercial classes as well as in popping beans or Nuñas. These seed types contrast with the elongated, cylindrical to kidney-shaped grain of race Nueva Granada genotypes (Voysest et al. 1994). It was interesting to note the none of the popping beans or ñuñas were found outside of race Peru and there was no evidence of introgression from other races or gene pools except for G23604 that appeared to be an intermediate between race Peru and race Chile. The diversity of popping beans adds to the diversity of race Peru in general and agree with results from Tohme et al. (1995b) who found a wide range of phaseolin alleles in popping beans, from T, being the most common, to C, H, Ca, Nu, A and Ko alleles. The fact that the popping beans showed little evidence of introgression from the Nueva Granada race does agree with the geographic range of popping beans, since these exist only as far north as Ecuador but are found as far south as Argentina (Tohme et al. 1995b).

In Colombia, we found that both the Nueva Granada and Peru races are present with distinctions between the two races validating the classification as discussed above, although the differences between the races here were less abrupt than between the races for the Andean race representatives. In particular, a continuum of admixture was observed between Nueva Granada and Peru races suggesting gene flow and a blurring of racial boundaries between them in Colombia. This was seen in the seed types of race Peru in Colombia that included both round and cylindrical shaped grain. Colombian race Nueva Granada genotypes could also be subdivided into subgroups suggesting that subraces might exist in this race as was found for the Andean race representatives. Gene flow was also evident between Andean and Mesoamerican gene pools with some Colombian genotypes with Andean type phaseolins showing similarity to the Mesoamerican control genotypes. Introgression between the gene pools has been amply reported and has acted to increase diversity within Andean beans both within the primary center of origin (Beebe et al. 2001; Islam et al. 2002) and in secondary centers of diversity (Santalla et al. 2002; Maciel et al. 2003). Introgression between races and between gene pools as well as potentially with wild accessions is the likely explanation for the high diversity observed in the Colombian germplasm. Colombia and northwestern South America generally have been cross-roads

For Peer Review

for common bean germplasm exchange since pre-colonial times due to its position at the landbridge with Central America and near the terminus of the Andes mountains. A diversity of wild accessions from both the Mesoamerican and Andean gene pool also exists in this region with weedy intermediates between wild and cultivated types (Debouck et al. 1993; Beebe et al. 1997). An early study by Gepts and Bliss (1986) identified S phaseolin along with the CH allele from the wild, and that cultivated genotypes in addition to having S, T and C phaseolin also had a the phaseolin allele B with the presence of this allele suggesting the possibility of additional domestication in this region (Gepts et al. 1986).

Now that race structure has been so clearly detected in Andean beans an open question is how this diversity arose in the gene pool. It is clear that the three Andean races are very distinct based on adaptive range, geographical distribution, morphological and agronomic traits, however they are suspected to have arisen from a single domestication event in the southern part of the range (Beebe et al. 2001; Chacón et al. 2005) therefore the source of race structure for the Andean gene pool is still unclear. It is evident however, that there is a need to change the paradigm for Andean races from one based on agro-morphological distinctions to one that is based on diversity assessment as was done here. Our results suggest that Andean races are separated mainly on the basis of seed types and ecological adaptation rather than growth habit alone. This understanding could lead to more rational use of germplasm and we expect that our results will aid in the design of crossing programs for Andean beans. Given their large seed size, bright seed colors and other unique characteristics in terms of seed color pattern and popping ability, Andean beans include many specialty types that often command high prices and as such are a valuable alternative to farmers in both developing and developed countries. Our results showing two major subdivisions to the gene pool that are based more on agroecological distinctions than on morphological differences indicate that crossing programs should emphasize crosses between groups defined in this manner to maximize genetic diversity. Another useful observation of our study is that more diversity is found for the nuclear genome of Andean beans than previously thought based on analysis of AFLP patterns and chloroplast diversity (Beebe et al. 2001; Chacón et al. 2005).

For Peer Review

In conclusion, our results will allow future studies to classify genotypes into Andean races more easily and will open the way for more detailed genetic mapping experiments in Andean beans. The characterization of genetic diversity within the Andean gene pool is useful for gene bank management, the development of reference germplasm collections for association studies and for preservation of landraces that are threatened in all three major races. For all these reasons the analysis of race structure in Andean beans is important and justifies the detailed analysis with microsatellite markers described in this research.

Acknowledgements: The authors wish to thank S. Beebe and C. de Vicente for helpful suggestions during the study. We are also grateful to D. Debouck, O. Toro and A. Hoyos of the Genetic Resource Unit and Bean Projects of CIAT for germplasm. This research was part of theses conducted at Univ. Nacional de Colombia and was supported by the Generation Challenge Program and CIAT funds.

For Peer Review

Figures

Figure 1. Multiple correspondence analysis showing spatial distribution of Chile (diamonds) Nueva Granada (hearts), and Peru (stars) race representatives according to the classification of Singh et al. (1991a). Mesoamerican and Andean control genotypes (circles) are indicated along with group showing introgression between the gene pools.

Figure 2. Dendrograms showing the relationship between a) Andean race representatives of common bean and b) Colombian genotypes based on Euclidean distance obtained by multiple correspondence analysis.

Figure 3. Prevalence index for genotypes belonging to Andean race representatives (A) and Colombian germplasm (B). Group abbreviations are C = race Chile; NG = race Nueva Granada; NG-A = subrgroup A of race Nueva Granada; NG-B = subgroup B of race Nueva Granada (see text); P = race Peru. Genotypes showing introgression (I) and those from the Mesoamerican gene pool (M) are shown for Andean race representatives.

Figure 4. Graphs of population structure (at K=3 and K=4) for a) Andean race representatives of common bean and b) Colombian genotypes, organized according to dendrogram clustering on the x-axes. Membership coefficients (y-axis) within clusters were determined based on 30,000 iterations using the software program Structure (Pritchard et al., 2000). Clusters are separated by vertical lines with cluster names indicated above figures (CH = race Chile; NG = race Nueva Granada; P = race Peru; INT = introgression with Mesoamerican genotypes; M = Mesoamerican control genotypes).

For Peer Review

References

Afanador L, Hadley S, Kelly JD (1993) Adoption of a mini-prep DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L). Bean Improv Coop 36:10–11

Becerra V, Gepts P (1994) RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. Genome 37:256–263

Beebe S, Toro O, González AV, Chacón MI, Debouck D (1997) Wild-weed–crop complex of common bean (*Phaseolus vulgaris* L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. Genet Resour Crop Evol 44:73–91

Beebe S, Skroch P, Tohme J, Duque MC, Pedraza F, Nienhuis J (2000) Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. Crop Sci 40:264–273

Beebe S, Renjifo J, Gaitán-Solís E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. Crop Sci 41:854–862

Beer S, Siripoonwiwat W, O'Donoughue L, Souza E, Matthews D, Sorrells M (1997) Associations between molecular markers and quantitative traits in an oat germplasm pool: Can we infer linkages? J Agric Genomics 3:1–16

Blair MW, Pedraza F, Buendia H, Gaitan E, Beebe S, Gepts P, Tohme J (2003) Development of a genome wide anchored microsatellite for common bean (*Phaseolus vulgaris* L). Theor Appl Genet 107:1362–1374

Blair MW, Giraldo MC, Buendia HF, Tovar E, Duque MC, Beebe S (2006) Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 113:100–109

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Bean (*Phaseolus* spp.); model food legumes. Plant Soil 252:55–128

Chacón MI, Pickersgill S, Debouck D (2005) Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of Mesoamerican and Andean cultivated races. Theor Appl Genet 110:432–444

Debouck D, Toro O, Paredes O, Johnson W, Gepts P (1993) Genetic diversity and ecological distribution of *Phaseolus vulgaris* L. (Fabacea) in northwestern south America. Econ Bot 47:408-423

Díaz LM, Blair MW (2006) Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. Theor Appl Genet 114: 143-54

Diaz LM, Diaz JM, Blair MW (2005) Diversidad genetica de frijol comun (*Phaseolus vulgaris L)* en Colombia. Fitotecnia Colombiana 5:28-36

Duarte J, Dos Santos J, Melo L (1999) Genetic divergence among common beans cultivars from different races based on RAPD markers. Genet Mol Biol 22:419–426

Durán LA, Blair MW, Giraldo MC, Machiavelli R, Prophete E, Nin JC, Beaver JS (2005) Morphological and molecular characterization of common bean (*Phaseolus vulgaris* L) landraces from the Caribbean. Crop Sci 45:1320-1328

For Peer Review

Gaitán E, Duque MC, Edwards K, Tohme J (2002) Microsatellite repeats in common bean (*Phaseolus vulgaris* L.): isolation, characterization, and cross-species amplification in *Phaseolus* spp. Crop Sci 42:2128–2136

Gepts P, Osborn T, Rashka K, Bliss F (1986) Phaseolin–protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris* L.): evidence for multiple centers of domestication. Econ Bot 40:451–468

Gepts P, Bliss F (1986) Phaseolin variability among wild and cultivated common bean (*Phaseolus vulgaris* L.) from Colombia. Econ Bot 40:469-478.

Gepts P (1988) Phaseolin as an evolutionary marker. In: Resources of *Phaseolus* beans. Kluwer, Dordtrecht, pp 215–241

Gepts P (1998) Origin and evolution of common bean: past events and recent trends. HortScience 33:1124-1130

Islam FM, Beebe S, Muñoz M, Tohme J, Redden RJ, Basford KE (2004) Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. Theor Appl Genet 108:243–252

Islam FM, Basford KE, Redden RJ, Gonzalez AV, Kroonenberg PM, Beebe SE (2002) Genetic variability in cultivated common bean beyond the two major gene pools. Genet Resour Crop Evol 49:271–283

Johns M, Skroch P, Nienhuis J, Hinrichsen P, Bascur G, Muñoz-Schick C (1997) Gene pool classification of common bean landraces from Chile based on RAPD and Morphological data. Crop Sci 37:605-613

Koenig R, Gepts P (1989) Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two mayor centers of genetic diversity. Theor Appl Genet 78:809–817

Maciel F, Echeverrigaray S, Gerald L, Grazziotin F (2003) Genetic relationship and diversity among Brazilian cultivar and of common beans (*Phaseolus vulgaris* L.) revealed by AFLP markers. Genet Resour Crop Evol 50:887-893

Metais I, Hamon B, Jalouzot R, Peltier D (2002) Structure and level of genetic diversity in various bean types evidenced with microsatellite markers isolated from a genomic enriched library. Theor Appl Genet 104:1346–1352

Nei M (1987) Molecular evolutionary genetics. Columbia University press, New York.

Nei M (1978) Estimation of average heterozygocity and genetic distance from a small number of individuals. Genetics 89:583–590

Paredes O, Gepts P (1995) Extensive introgression of middle American germplasm into Chilean common bean cultivars. Genet Resour Crop Evol 42:29–41

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Rohlf F (2002) NTSYS pc. Numerical Taxonomy System Exeter Publishing, Setauket, NY

Rosenberg NA (2002) Distruct: a program for the graphical display of structure results. http://www.cmb.usc.edu/»noahr/distruct.

Santalla M, Rodiño A, de Ron A (2002) Allozyme evidence supporting southwestern Europe as a secondary center of genetic diversity for the common bean. Theor Appl Genet 104:934–944

Santalla M, Menendez-Sevillano M, Monteagudo A, de Ron A (2004) Genetic diversity of Argentinian common bean and its evolution during domestication. Euphytica 135:75-87

SAS-Institute (1996) SAS/STAT Users guide 6.11. SAS Inst, Cary, NC.

For Peer Review

Singh S (1989) Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 43:39–57

Singh S, Gepts P, Debouck D (1991a) Races of common bean (*Phaseolus vulgaris*, Fabaceae). Econ Bot 45:379–96

Singh S, Gutierrez A, Molina A, Urrea C, Gepts P (1991b) Genetic diversity in cultivated common bean. II. Marker-based analysis on morphological and agronomic traits. Crop Sci 31:23–29

Singh S, Nodari R, Gepts P (1991c) Genetic diversity in cultivated common bean. I. Allozymes. Crop Sci 31:19–23

Slatkin M, Barton N (1989) A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349-1368

Sonnante G, Stockton T, Nodari R, Becerra Velasquez V, Gepts P (1994) Evolution of genetic diversity during the domestication of common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 89:629-635

Tohme J, Jones P, Beebe S, Iwanaga M (1995a) The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection, In Hodgkin, T., et al., eds. Core collections of plant genetic resources. Wiley Press

Tohme J, Toro O, Vargas J, Debouck D (1995b) Variability in Andean ñuña common beans (*Phaseolus vulgaris*, Fabaceae). Econ Bot 49:78-95

Voysest O, Valencia M, Amezquita M (1994) Genetic diversity among Latin American Andean and Mesoamerican common bean cultivars. Crop Sci 34:1100–1110

Yeh FY, Boyle R, Ye T, Mao Z (1997) POPGENE, the user friendly shareware for population genetic analysis, version 1.31. Molecular Biology and Biotechnology Centre, University of Alberta, Canada. Molecular Biology and Biotechnology Centre

Yu K, Park J, Poysa V, Gepts P (2000) Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). J Hered 91:429–434

For Peer Review

Table 1. Genotypes evaluated in the study of Andean race diversity.

Genotype	Common Name	Country of Origin (Department or State)	Growth Habit	Seed Size	Phaseolin	Status
Andean Race Representatives						
Race Nueva Granada						
G1326	Canario	Mexico (Guanajuato)	I	Medium	T	Landrace
G1448	Saxa	Yugoslavia	I	Medium	T	Commercial
G2488	Canario	Mexico (Michoacan)	I	Large	T	Landrace
G4080	Radical	Colombia (Valle de Cauca)	I	Large	T	Landrace
G4452	ICA Guali	Colombia	I	Large	T	Commercial
G4494	Diacol Calima	Colombia (Valle de Cauca)	I	Large	T	Commercial
G4520	Estrada Rosado	Colombia (Antioquia)	I	Large	T	Landrace
G4681	Guarzo	Colombia (Nariño)	I	Large	CA	Landrace
G4717	Radical	Colombia (Valle de Cauca)	I	Large	T	Landrace
G4906	Radical	Colombia (Valle de Cauca)	I	Large	CA	Landrace
G5024	Jalo	Brazil	III	Medium	T	Landrace
G5254	Bagajo	Brazil	III	Medium	T	Landrace
G5708	Sangretoro	Colombia (Antioquia)	I	Large	T	Landrace
G6379	ICA Tone	Colombia (Antioquia)	I	Large	T	Commercial
G7930	Alubia (Sel. Cerrillos)	Argentina (Salta)	I	Medium	T	Commercial
G8147	Uribe Redondo	Colombia (Antioquia)	I	Large	T	Landrace
G9603	Jalo EEP 558	Brazil (Minas Gerais)	III	Medium	T	Landrace
G12368	Canario Alargado	Ecuador (Loja)	I	Medium	T	Landrace
G12720	Calabozo	Colombia (Antioquia)	I	Medium	C	Landrace
G14660	Uribe	Colombia	I	Large	T	Landrace
G18255	Velazco Largo	Cuba	I	Large	T	Commercial
G18264	Pompadour Checa 50	Dom. Rep.	I	Medium	T	Commercial
G21720	Cargabello	Ecuador	I	Medium	T	Commercial
G21953	Jalo	Brazil	III	Medium	T	Landrace
G21227	Mortiño Alargado	Colombia (Nariño)	I	Medium	T	Landrace
Race Chile						
G2544	Frutilla	Chile	I	Large	NA	Landrace

For Peer Review

G4472	Tortolas	Chile	III	Large	TO	Landrace
G4474	Coscorron	Chile	III	Large	NA	Landrace
G5831	Ganso	Chile	III	Large	PA	Landrace
G18356	Frutilla	Chile	I	Large	T	Commercial
G18372	Coscorron	Chile	I	Large	T	Commercial
Race Peru						
G111	Huasca Huallaga Colorado	Peru (Amazonas)	IV	Medium	T	Landrace
G5700	Bayo Bolon	Ecuador (Guayas)	IV	Large	T	Landrace
G5702	Cargamanto	Colombia (Antioquia)	IV	Large	H1	Landrace
G5723	Caballero	Peru	IV	Medium	T	Landrace
G7231	Cargamanto	Colombia (Antioquia)	IV	Large	H	Landrace
G7309	Sangretoro	Colombia (Cundinamarca)	IV	Large	T	Landrace
G7314	Huevo de Pinche	Colombia (Caldas)	IV	Small	H	Landrace
G11785	Nuna	Peru (Huanuco)	III	Large	Pa-T	Landrace
G11819	Liborino Voluble	Colombia (Antioquia)	IV	Large	T	Landrace
G12207	Canario	Ecuador (Chimborazo)	IV	Large	C	Landrace
G12209	Bolon Rojo	Ecuador (Chimborazo)	IV	Large	C	Landrace
G12229	Bola	Ecuador (Asuay)	IV	Large	C	Landrace
G12327	Cargamanto	Ecuador	IV	Large	T-H1	Landrace
G12407	Bolon Bayo	Ecuador (Imbabura)	IV	Large	T	Landrace
G12421	Bolon Rojo	Ecuador (Guayas)	IV	Large	C	Landrace
G12438	Bolon Amarillo-1	Ecuador (Guayas)	IV	Large	C	Landrace
G12572	Nuña Mani Palida-1	Peru (Cajamarca)	IV	Large	H	Popping bean
G12587	Nuña Callashina Oscura	Peru (Cajamarca)	IV	Large	H	Popping bean
G12621	Nuña	Peru (Ancash)	IV	Medium	T	Popping bean
G12623	Numia Plana	Peru (Ancash)	IV	Large	T	Popping bean
G12644	Morado	Colombia (Antioquia)	IV	Large	H	Landrace
G12709	Mortino	Colombia (Nariño)	IV	Large	H	Landrace
G13940	Overito	Argentina (Salta)	IV	Large	T	Landrace
G18325	Caballero	Peru	III	Large	T	Landrace
G19833	Chaucha Chuga	Peru (Amazonas)	III	Large	H	Landrace
G23604	Nuña	Peru (Apurimac)	III-IV	Medium	T	Popping bean
G23614	Nuña	Peru (Cusco)	III-IV	Large	T	Popping bean
G23691	Nuña Cenizo	Peru (La Libertad)	IV	Medium	T	Popping bean

For Peer Review

Colombian Genotypes						
Landraces						
G4494	Diacol Calima	Colombia (Valle de Cauca)	I	Large	T	Commercial
G4547	Liborino De Mata	Colombia (Antioquia)	I	Large	H	Landrace
G4564	Boca De Angel	Colombia (Antioquia)	I	Large	CA	Landrace
G4613	Guarzo O Sesentano	Colombia (Cauca)	III	Large	CA	Landrace
G4644	Limoncillo	Colombia (Cundinamarca)	I	Large	T	Landrace
G4653	Cuarentano	Colombia (Huila)	IV	Large	T	Landrace
G4672	Santa Lucia	Colombia (Nariño)	III	Small	CA	Landrace
G4681	Guarzo	Colombia (Nariño)	I	Large	CA	Landrace
G4691	Matahambre	Colombia (Nariño)	III	Small	CH	Landrace
G4780	Revoltura	Colombia (Cauca)	III	Large	T	Landrace
G4906	Radical	Colombia (Valle de Cauca)	I	Large	CA	Landrace
G5708	Sangretoro	Colombia (Antioquia)	I	Large	T	Landrace
G7229	Culateño	Colombia (Caldas)	IV	Large	H	Landrace
G7231	Cargamanto	Colombia (Antioquia)	IV	Large	H	Landrace
G7257		Colombia (Boyacá)	IV	Medium	T	Landrace
G7309	Sangretoro	Colombia (Cundinamarca)	IV	Large	T	Landrace
G7314	Huevo De Pinche	Colombia (Caldas)	IV	Small	H	Landrace
G7317	Revoltura	Colombia (Antioquia)	IV	Large	H	Landrace
G7320	Chiguano	Colombia (Cundinamarca)	IV	Small	H	Landrace
G7381		Colombia (Cauca)	IV	Large	H	Landrace
G7385	Uribe Redondo	Colombia (Antioquia)	I	Medium	T	Landrace
G7437	Huevo De Pinche	Colombia (Cauca)	IV	Large	H	Landrace
G7440	Culateno	Colombia (Cauca)	IV	Large	T	Landrace
G8157	Revoltura	Colombia (Antioquia)	IV	Large	H	Landrace
G8160	Plomo	Colombia (Antioquia)	IV	Large	T	Landrace
G8208	Blanco Enredadera	Colombia (Cundinamarca)	IV	Medium	H	Landrace
G9542	Uribe Largo	Colombia (Antioquia)	I	Large	T	Landrace
G11819	Liborino Voluble	Colombia (Antioquia)	IV	Large	T	Landrace
G12648	Higuerillo	Colombia (Antioquia)	IV	Large	T	Landrace
G12657	Bala	Colombia (Cauca)	IV	Large	H	Landrace
G12661	Tena	Colombia (Cundinamarca)	IV	Large	H	Landrace
G12667	Blanco Sabanero	Colombia (Nariño)	IV	Large	H	Landrace
G12689	Mandarino	Colombia (Nariño)	I	Large	T	Landrace

For Peer Review

G12692	Guarzo	Colombia (Nariño)	IV	Medium	H	Landrace
G12702	Liborino Pintas Rojas	Colombia (Nariño)	IV	Large	T	Landrace
G12706	Liborino Pintas Negras	Colombia (Nariño)	IV	Large	T	Landrace
G12709	Mortiño	Colombia (Nariño)	IV	Large	H	Landrace
G12709A	Mortiño Rojo	Colombia (Nariño)	IV	Large	H	Landrace
G12714	Care Raton	Colombia (Nariño)	IV	Medium	T	Landrace
G12715		Colombia (Nariño)	IV	Large	T	Landrace
G13220		Colombia (Tolima)	I	Large	T	Landrace
G14016	ICA Tundama	Colombia (Antioquia)	III	Large	T	Commercial
G14643	Tunia 1	Colombia (Cauca)	IV	Large	CA	Landrace
G15839	Tierra Adentro 1	Colombia (Cauca)	IV	Large	T	Landrace
G16840		Colombia (Cauca)	IV	Large	TO	Landrace
G19142B	Criollo	Colombia (Cundinamarca)	IV	Large	CA	Landrace
G21202	De Guasca	Colombia (Nariño)	IV	Large	T	Landrace
G21210	Monte Oscuro	Colombia (Nariño)	I	Large	T	Landrace
G21217	Bola Mani	Colombia (Nariño)	I	Large	H	Landrace
G21227	Mortiño Alargado	Colombia (Nariño)	I	Medium	T	Landrace
G21242		Colombia (Nariño)	III	Medium	T	Landrace
G23682		Colombia (Antioquia)	IV	Large	T	Landrace
Advanced Lines						
ABA2		Colombia - CIAT	I	Medium	T	Advanced line
AND1005		Colombia - CIAT	II	Medium	T	Advanced line
BAT1373		Colombia - CIAT	I	Medium	T	Advanced line
CAL149		Colombia - CIAT	II	Large	T	Advanced line
DRK47		Colombia - CIAT	I	Large	T	Advanced line
LRK31		Colombia - CIAT	II	Large	T	Advanced line
PVA773		Colombia - CIAT	I	Large	T	Advanced line
PVA1111		Colombia - CIAT	I	Large	T	Advanced line
Control Genotypes						
G4494	Calima	Colombia - ICA	I	Large	T	Commercial
G19833	Chaucha Chuga	Peru (Amazonas)	III	Large	T	Landrace
DOR364	Dorado	Colombia - CIAT	II	Small	S	Advanced line
ICA PIJAO	Pijao	Colombia – ICA	II	Small	S	Commercial

For Peer Review

Table 2. Distribution of phaseolin types in the Andean race representatives and Colombian genotypes used in the study.

Study/Country of Origin	No. [1]	Andean Races																			
		Nueva Granada					Chile				Peru							Introgressed			
		T	C	CA	T/H1	H	T	Pa	To	NA[2]	H	T/H1	C	CA	T	Pa/T	NA[2]	CA	CH	H	T
Andean Race Representatives																					
Argentina	2	1					1														
Brazil	4	2																			2
Colombia	20	11	1	2	1	1					2						1			1	
Cuba	1	1																			
Chile	6	2						1	1	1							1				
Dominican Republic	1	1																			
Ecuador	10	2										1	5		2						
Mexico	2	2																			
Peru	12	1									3				5	1	1				1
Yugoslavia	1	1																			
Total	59	24	1	2	1	1	1	1	1	1	5	1	5	0	7	1	1	0	0	1	3
Colombian genotypes	60	23	0	5	0	10	-	-	-	-	4	0	0	1	11	0	0	1	1	2	2

1/ Control genotypes not shown 2/ NA = phaseolin pattern unidentified

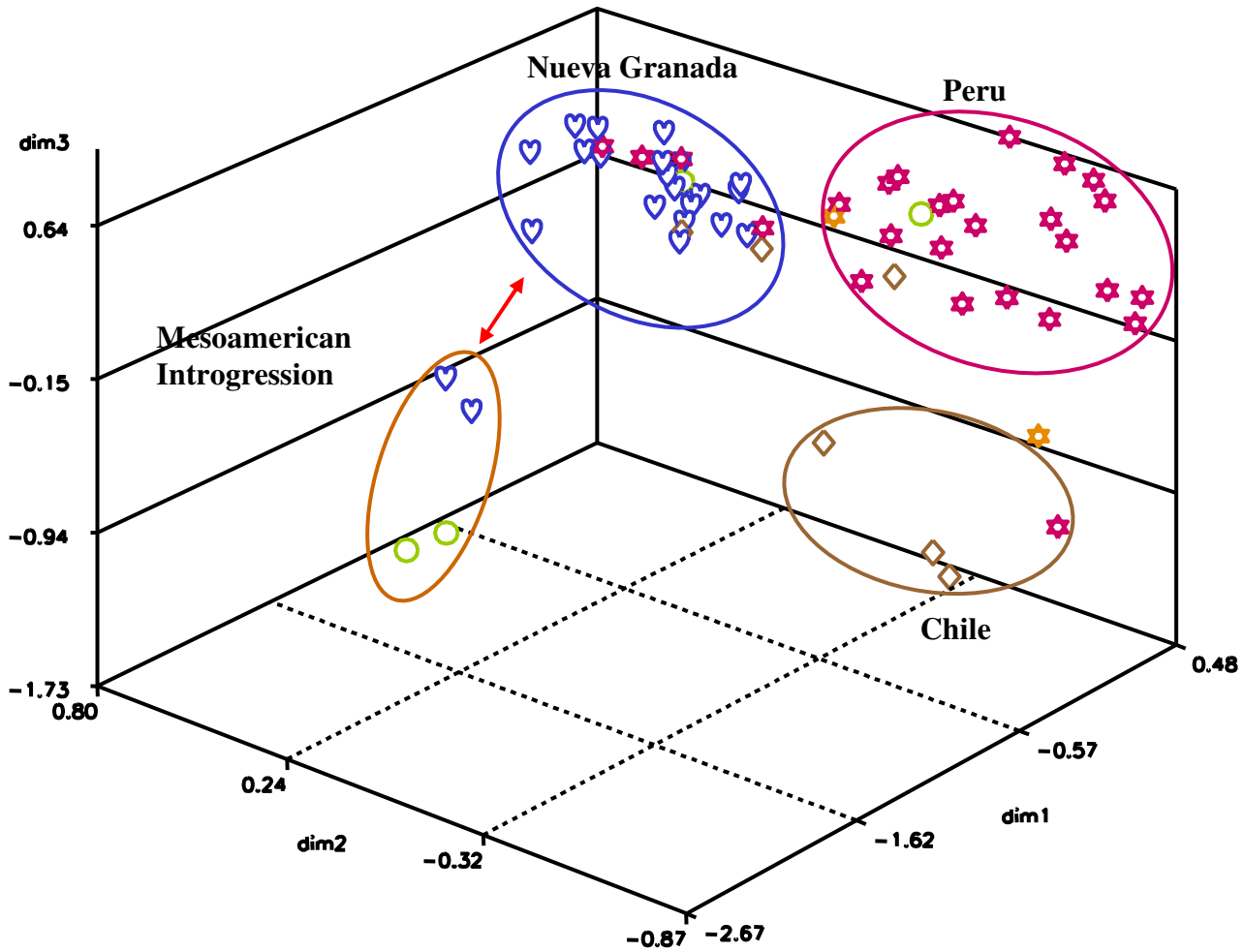
For Peer Review

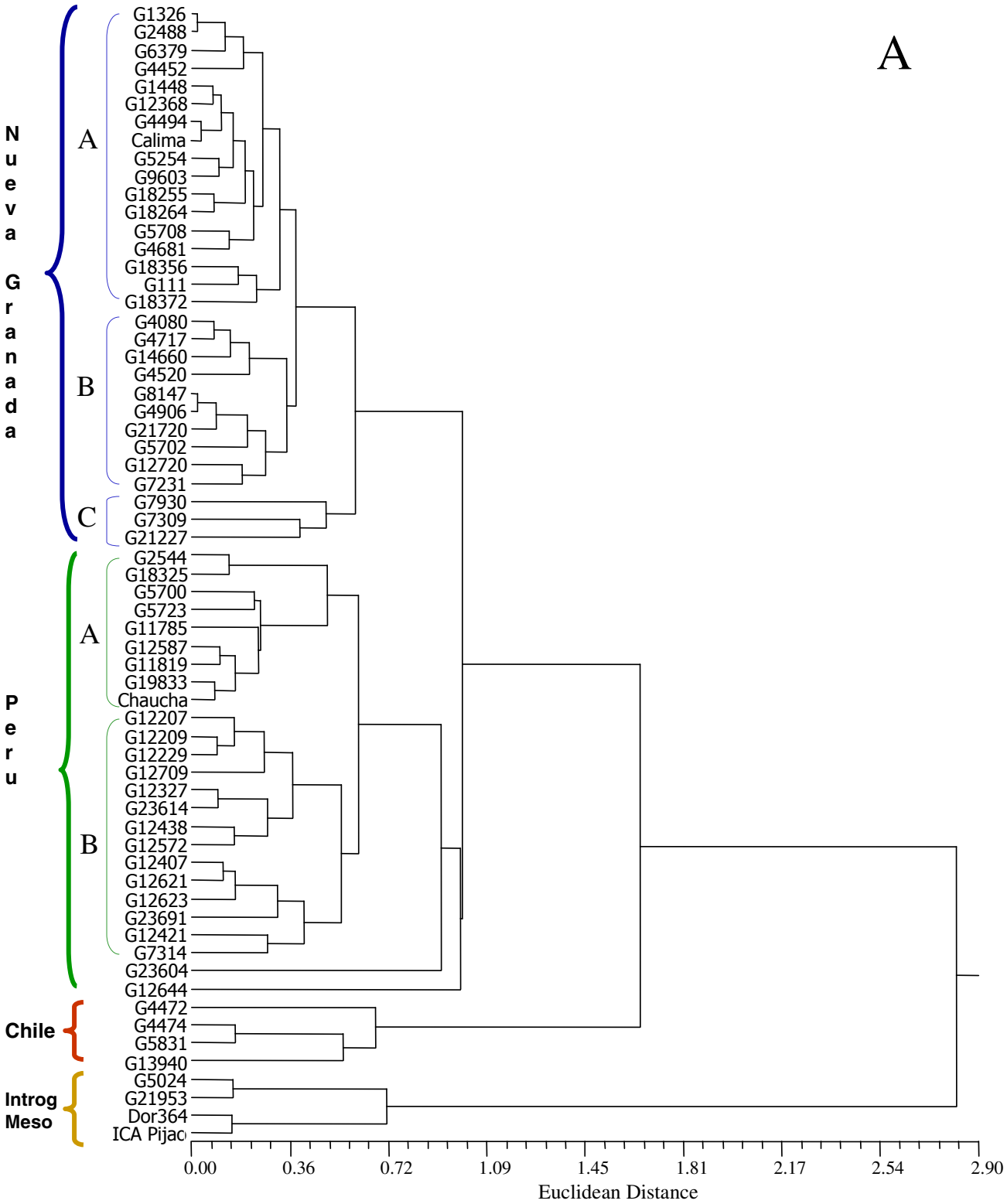
Table 3. Growth habit and seed size for the Andean race representatives.

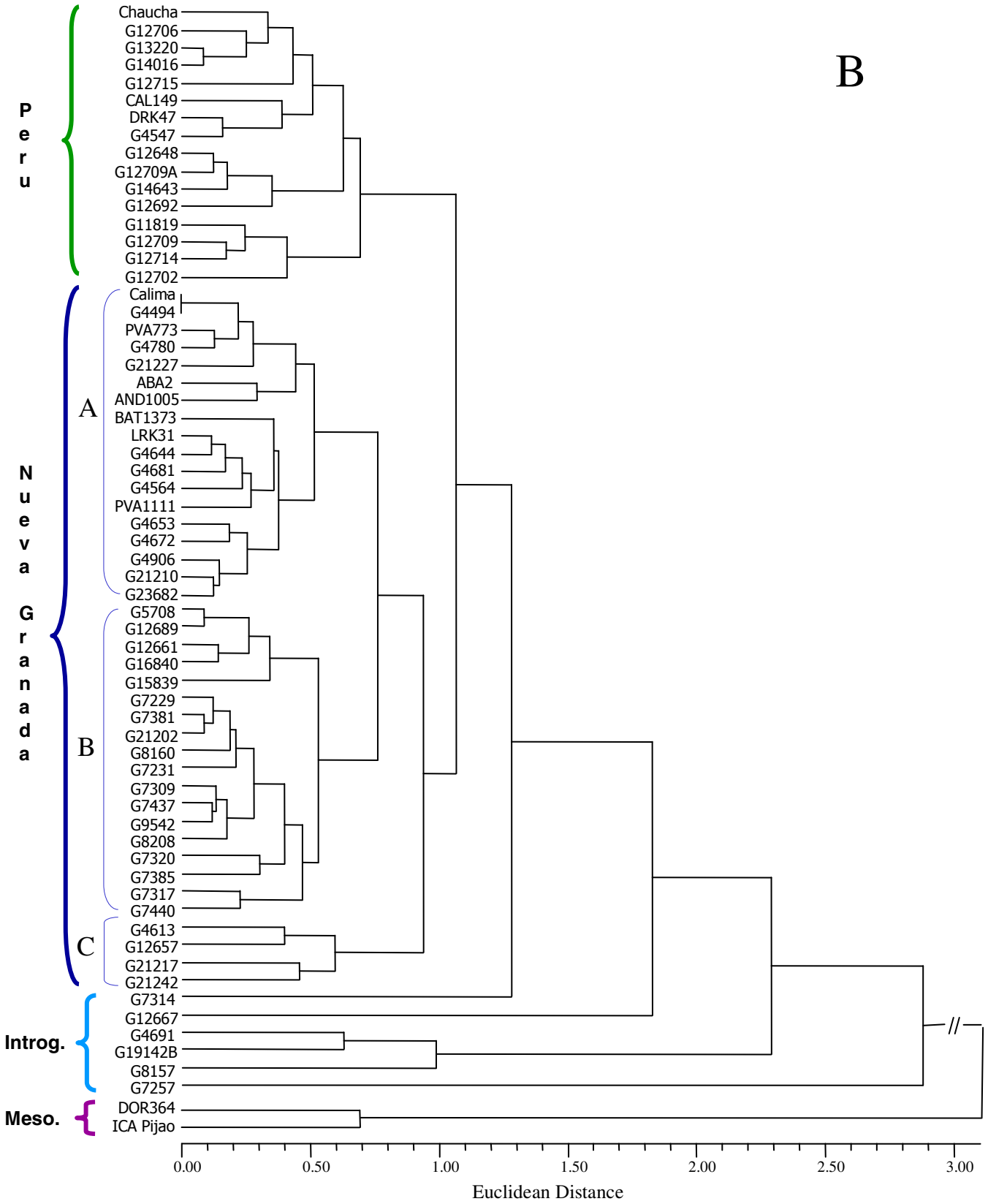
Race	No. [1]	Growth Habit					Seed Size		
		I	II	III	IV	NA [2]	Small	Medium	Large
Nueva Granada	30	23		2	4	1		11	19
Chile	4			3		1			4
Peru	25	1		3	18	1	1	4	20
Introgressed	4	1	1	2			3	1	
Total	63	24	0	8	22	3	1	15	41

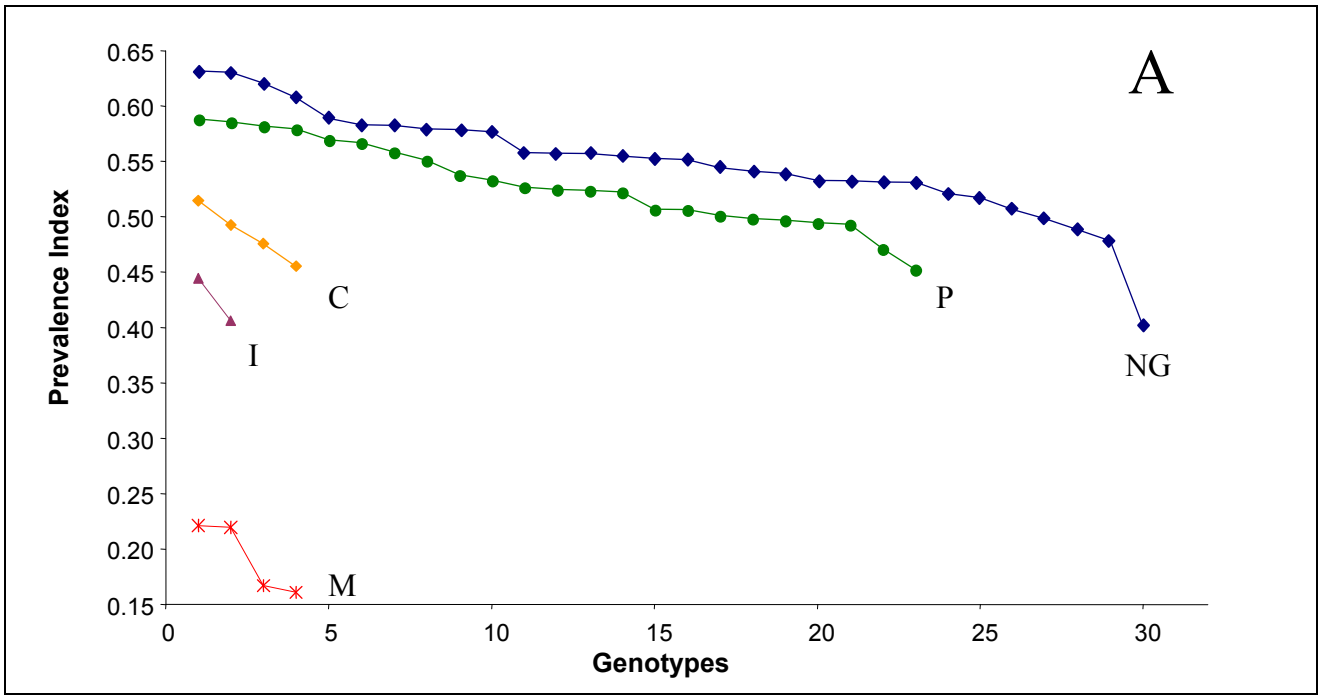
1/ Control genotypes included 2/ NA = growth habit not determined

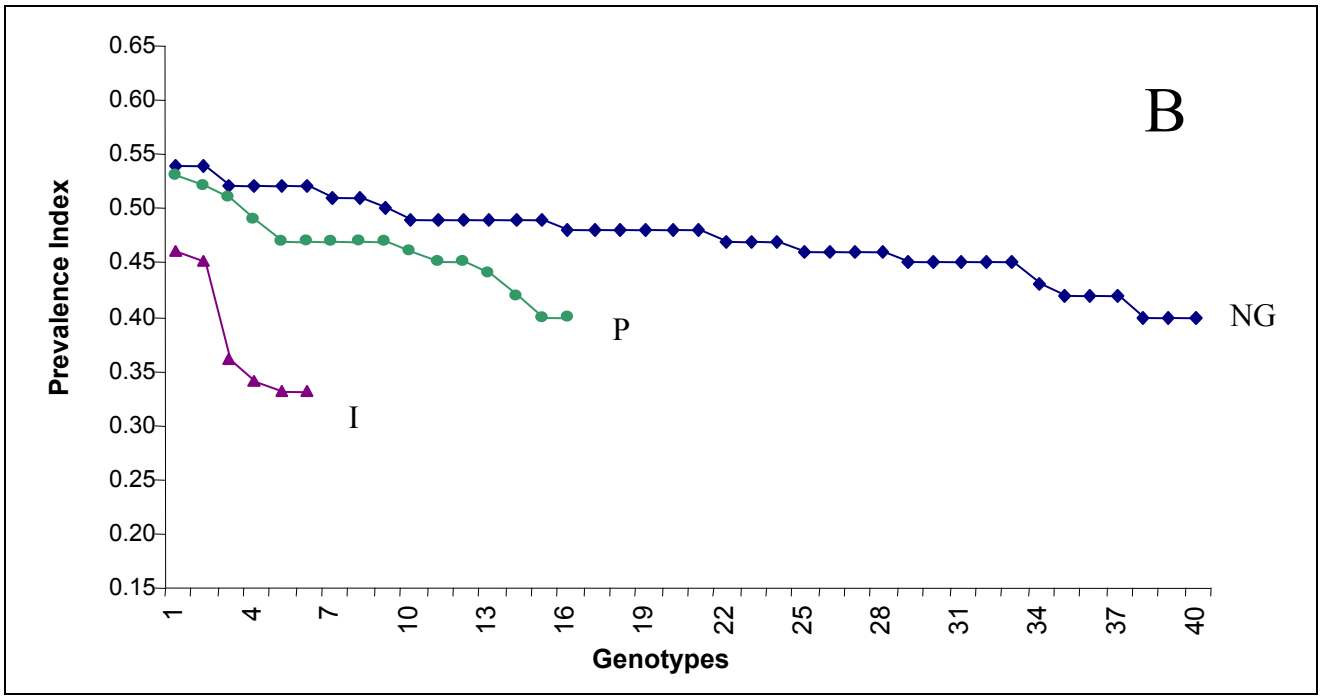
For Peer Review


Nueva Granada
Peru
Mesoamerican
Introgression
Chile
dim3
0.64
−0.15
−0.94
−1.73
dim2
0.80
0.24
−0.32
−0.87
dim1
0.48
−0.57
−1.62
−2.67

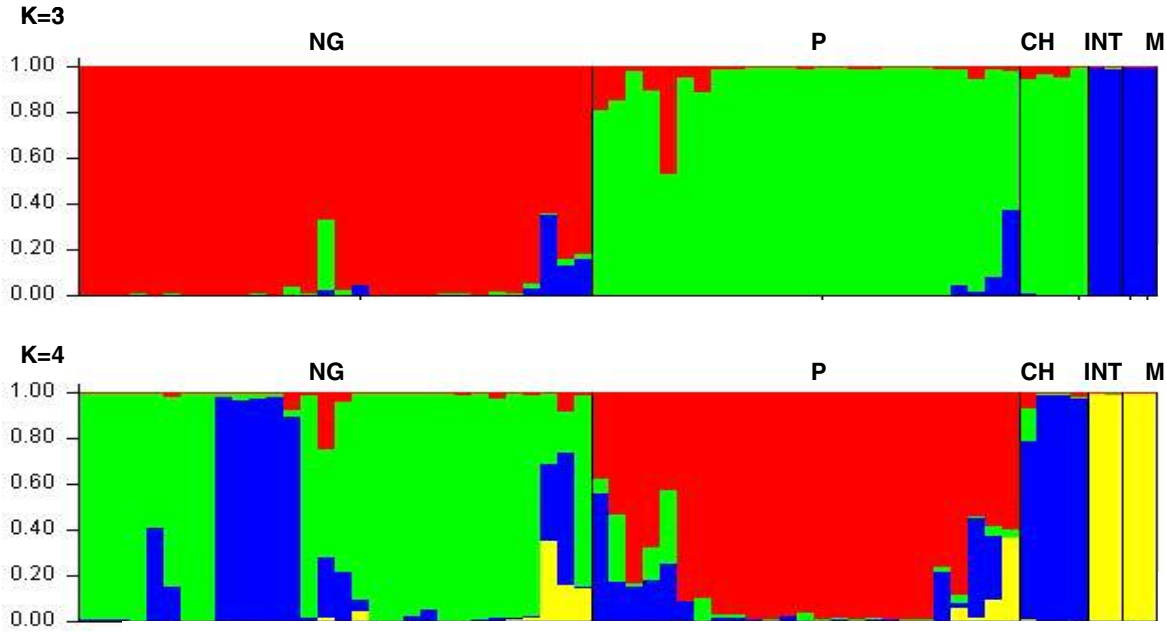

A
Nueva Granada
Peru
Chile
Introg Meso
A
B
C
A
B
G1326
G2488
G6379
G4452
G1448
G12368
G4494
Calima
G5254
G9603
G18255
G18264
G5708
G4681
G18356
G111
G18372
G4080
G4717
G14660
G4520
G8147
G4906
G21720
G5702
G12720
G7231
G7930
G7309
G21227
G2544
G18325
G5700
G5723
G11785
G12587
G11819
G19833
Chaucha
G12207
G12209
G12229
G12709
G12327
G23614
G12438
G12572
G12407
G12621
G12623
G23691
G12421
G7314
G23604
G12644
G4472
G4474
G5831
G13940
G5024
G21953
Dor364
ICA Pijao
0.00
0.36
0.72
1.09
1.45
1.81
2.17
2.54
2.90
Euclidean Distance


B
Peru
Nueva Granada
Introg.
Meso.
A
B
C
Chaucha
G12706
G13220
G14016
G12715
CAL149
DRK47
G4547
G12648
G12709A
G14643
G12692
G11819
G12709
G12714
G12702
Calima
G4494
PVA773
G4780
G21227
ABA2
AND1005
BAT1373
LRK31
G4644
G4681
G4564
PVA1111
G4653
G4672
G4906
G21210
G23682
G5708
G12689
G12661
G16840
G15839
G7229
G7381
G21202
G8160
G7231
G7309
G7437
G9542
G8208
G7320
G7385
G7317
G7440
G4613
G12657
G21217
G21242
G7314
G12667
G4691
G19142B
G8157
G7257
DOR364
ICA Pijao
0.00
0.50
1.00
1.50
2.00
2.50
3.00
Euclidean Distance

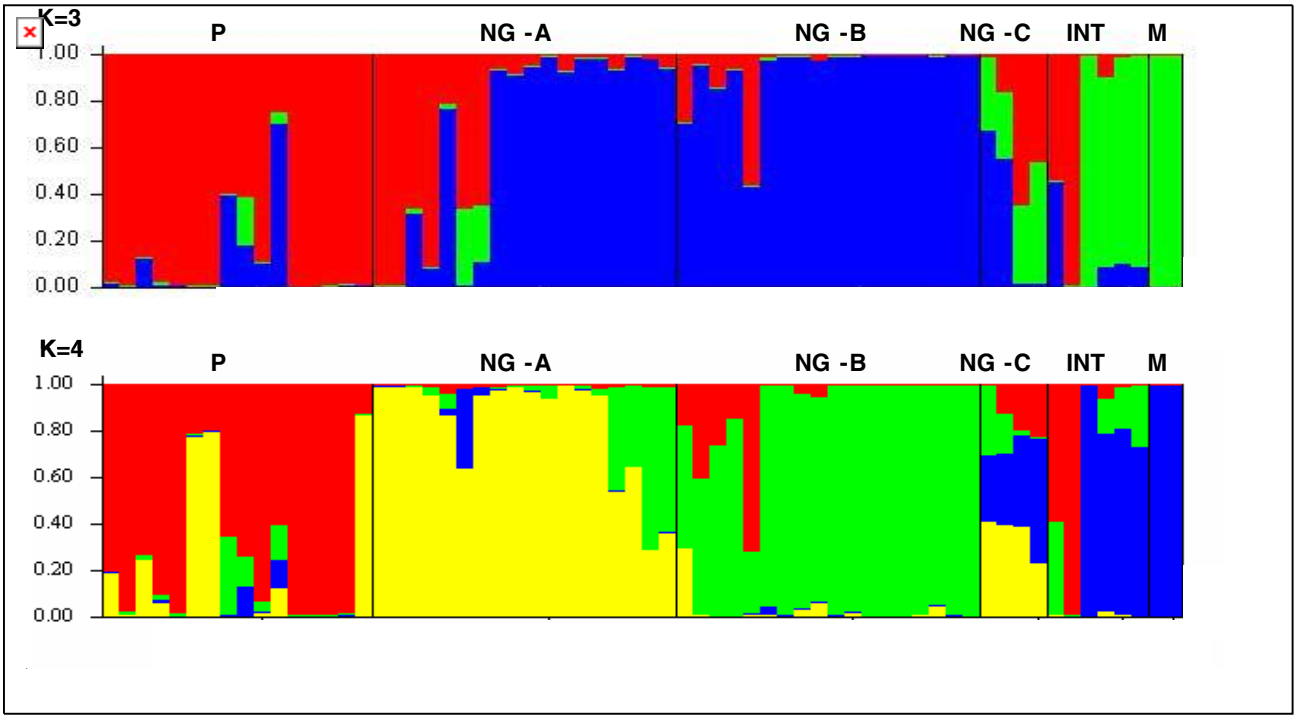

A
Prevalence Index
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0
5
10
15
20
25
30
Genotypes
C
P
I
NG
M


B
Prevalence Index
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
1
4
7
10
13
16
19
22
25
28
31
34
37
40
Genotypes
P
NG
I

A)


K=3
NG
P
CH
INT
M
1.00
0.80
0.60
0.40
0.20
0.00
K=4
NG
P
CH
INT
M
1.00
0.80
0.60
0.40
0.20
0.00

B)


K=3
P
NG -A
NG -B
NG -C
INT
M
1.00
0.80
0.60
0.40
0.20
0.00
K=4
P
NG -A
NG -B
NG -C
INT
M
1.00
0.80
0.60
0.40
0.20
0.00