Evaluating Two-Dimensional Electrophoresis Profiles of the Protein Phaseolin as Markers of Genetic Differentiation and Seed Protein Quality in Common Bean (Phaseolus vulgaris L.)

María López-Pedrouso[a], Javier Bernal[a], Daniel Franco[b], and Carlos Zapata[a*]

[a]Department of Genetics, University of Santiago de Compostela, Santiago de Compostela-15782, Spain

[b]Meat Technological Centre of Galicia, r/Galicia 4, Parque Tecnolóxico de Galícia, San Cibrán das Viñas, Ourense-32900, Spain

*corresponding autor:

Abstract:

High-resolution two-dimensional electrophoresis (2-DE) profiles of the protein phaseolin, the major seed storage protein of common bean, display great number of spots with differentially glycosylated and phosphorylated α- and β-type polypeptides. This work aims to test whether these complex profiles can be useful markers of genetic differentiation and seed protein quality in bean populations. The 2-DE phaseolin profile and the amino acid composition were examined in bean seeds from 18 domesticated and wild accessions belonging to the Mesoamerican and Andean gene pools. We found that proteomic distances based on 2-DE profiles were successful in identifying the accessions belonging to each gene pool and outliers distantly related. In addition, accessions identified as outliers from proteomic distances showed the highest levels of methionine content, an essential amino acid deficient in bean seeds. These findings suggest that 2-DE phaseolin profiles provide valuable information with potential of being used in common bean genetic improvement

Keywords: Phaseolus vulgaris, phaseolin, seed storage proteins, proteomic distance, amino acid composition, wild and cultivated beans

1. Introduction

The common bean (Phaseolus vulgaris L.) is the most important grain legume worldwide for direct human consumption, mainly in South America and Africa where it is a major dietary protein source.[1] In 2011, total world production of dry and green beans was about 23.3 and 20.4 million metric tons, respectively, harvested at 30.8 million hectares.[2] As a food, common bean seeds have valuable nutritional and functional properties. Thus, dry bean seeds are an important source of proteins, saccharides, and micronutrients, and they are rich in dietary fiber and low in fat. Besides their nutritive value, a regular intake of beans reduces common diseases such as coronary heart disease, type-II diabetes, and cancer. Nevertheless, the nutritional value of common bean seeds is affected by a number of factors, including low levels of sulfur-containing amino acids (methionine and cysteine) and tryptophan, low protein digestibility, and the presence of distinct antinutritional compounds.[3−6] In particular, the fraction of the essential amino acid methionine in bean seed protein is especially low as compared to the reference amino acid FAO/WHO requirements recommended in human nutrition.[5,7]

Much of our knowledge about the genetic differentiation and structure of wild and domesticated bean populations rests on the study of the variability of the seed protein phaseolin. The 7S globulin phaseolin is encoded by a gene family constituted by approximately 6−10 codominant genes per haploid genome, organized in a single close cluster of related genes on chromosome 7.[8,9] Members of the phaseolin gene family can be subdivided into two distinct gene subfamilies termed α and β encoding for α- and β-type polypeptides, respectively.[8] The one-dimensional sodium dodecyl sulfate-polyacrylamide gel

electrophoresis (1-DE) has been instrumental in the characterization of the variability of the protein phaseolin in common bean populations.[10−14] Thus, phaseolin diversity data based on 1-DE banding patterns have shown that domesticated common bean resulted from two already diverged major ecogeographic gene pools of wild P. vulgaris populations distributed from northern Mexico to Colombia (Mesoamerican gene pool) and from southern Peru to northwestern Argentina (Andean gene pool). The occurrence of domestication events in the two already diverged major gene pools of wild common bean populations resulted in the Mesoamerican and Andean domesticated gene pools. Additional evidence for these two gene pools was provided by morphological and phytopathological data along with different types of molecular markers including allozymes, lethality genes, and a diversity of nuclear, mitochondrial, and chloroplast DNA polymorphisms.[10,14−18] Following domestication, gene pool divergence led to the appearance of four Mesoamerican (Durango, Jalisco, Mesoamerica, and Guatemala) and three Andean (Nueva Granada, Peru, and Chile) races.[19]

The study of the phaseolin also provides useful information about important agronomic traits and biochemical characteristics of common bean cultivars. It has been found that types of phaseolin identified by 1-DE are consistently associated with seed weight across a broad range of genetic backgrounds and environments, tolerance to low soil pH, growth habit, early maturity, and resistance to parasites.[11,20,21] Moreover, the phaseolin is the major seed storage protein that may account for up to ca. 50% of the total seed proteome.[13] It is therefore largely responsible for the seed protein quality. In particular, the levels of phaseolin were found to be positively correlated with the nutritional availability

of methionine in total bean seed protein.[22] It has been suggested, therefore, that an improvement of the seed protein quality of the present bean varieties might be accomplished by using types of phaseolin with differences in the number of methionine residues and molar ratio of methionine.23

So far, more than 40 different types of phaseolin have been found in domesticated and wild cultivars of beans by 1-DE.[4,5] Nevertheless, 1-DE phaseolin profiles typically have only two to six bands with molecular mass (Mr) ranging from 40 to 54 kDa.[4,24] More informative reference maps of the protein phaseolin have been obtained only recently using high resolution two-dimensional electrophoresis (2-DE).[25,26] Application of high-resolution 2-DE provides characteristic and reproducible phaseolin profiles for each common bean population, each one constituted by a large number of spots containing differentially glycosylated and phosphorylated α-and β-type polypeptides (see Supporting Information Figures 1 and 2). This surge of information on the variation of the protein phaseolin enables one to obtain quantitative estimates of its degree of diversity across common bean populations based on measures of proteomic distance.[27]

The aim of the present work is to test whether highresolution 2-DE phaseolin profiles can be useful as molecular markers of the degree of genetic divergence and/or the quality of seed protein in common bean populations. For this purpose, we performed a prospective study on the variation of 2-DE phaseolin profiles over domesticated and wild common bean populations belonging to the Mesoamerican and Andean gene pools. In addition, the quality of the total seed protein in these bean populations was estimated from the content of methionine, an essential amino acid deficient in bean seeds.

From the results, we will try to determine whether proteomic distances based on 2-DE phaseolin profiles can efficiently identify populations belonging to each major gene pool, as well as the possible link between proteomic distances and seed protein quality. It might open the way of using the information contained in 2-DE phaseolin profiles for the genetic improvement of common bean populations.

2. Material and Methods

Plant Material. Dry mature seeds from 8 wild types and 10 cultivars of common bean (P. vulgaris L.) from Mesoamerican and Andean gene pools were screened for variability of phaseolin using 2-DE. The cultivars belong to the Mesoamerica, Nueva Granada, Peru, and Chile races. A complete list of the accessions used in this studied together with their passport information, which includes accession identifier, common name, phaseolin type assessed by 1-DE, race type, status (wild, landrace, or cultivar), origin (country and locality), seed weight, and seed color is given in Table 1. The seeds were provided by the United States Department of Agriculture (USDA). Ten bean seeds from each accession were lyophilized after previous grinding of the total seed tissue to a powder with liquid nitrogen using a precooled mortar and pestle. The samples were stored at −80 °C until protein extraction.

Seed Protein Extraction. Total seed protein was extracted by the phenol method, which has been proved to give highly satisfactory results for the separation of common bean seed proteins by 2-DE in comparison with other commonly used extraction methods.29 The phenol method was performed from 100 mg of sample suspended in extraction buffer (500 mM Tris-HCl, 500 mM EDTA, 700 mM sacarose, 100 mM KCl pH = 8.0, 2% DTT, and 1 mM PMSF).

Phenol saturated with Tris-HCl (pH 6.6−7.9) was added, and the phenolic phase was recovered after centrifuging (4500 rpm at 4 °C). Proteins were precipitated (0.1 M ammonium acetate in MeOH), and protein pellets were washed (0.1 M ammonium acetate, 10 mM DTT, 80% acetone, and 10 mM DTT) and resuspended in lysis buffer (7 M urea; 2 M thiourea; 4% CHAPS; 10 mM DTT, and 2% Pharmalyte pH 3−10, GE Healthcare, Uppsala, Sweden). Protein quantification was assessed for each extraction using the commercial CB-X protein assay kit (G-Biosciences, St. Louis, MO) following the instructions of the manufacturer for using a microplate reader.

2-DE Phaseolin Patterns. Optimized 2-DE protocol for highresolution phaseolin patterns was used as previously described.25,26 2-DE was carried out from 50 µg of total seed protein extract dissolved in lysis and rehydration (7 M urea, 2 M thiourea, 4% CHAPS, 0,002% Bromophenol Blue) buffers. Protein extracts were loaded on 24 cmlong ReadyStrip IPG Strips with a pH linear gradient of 4.7−5.9 (Bio-Rad Laboratories, Inc., Hercules, CA), together with 0.6% dithiothreitol (DTT) and 1% IPG buffer (Bio-Rad Laboratories).

First dimensional isoelectric focusing (IEF) was carried out until reaching 60 kVh after an initial rehydration step for 12 h at 50 V in a PROTEAN IEF cell (Bio-Rad Laboratories). Focused strips were incubated in the equilibration solution (50 mM Tris pH = 8.8, 6 M urea, 30% glycerol, and 2% SDS) with 1% DTT for 15 min at room temperature and then with 2.5% iodoacetamide under the same conditions. For the second dimension (SDS-PAGE), equilibrated strips were placed on 10% (w/v) gels, and they were run vertically in an Ettan DALTsix system (GE, Healthcare, Uppsala, Sweden) using Tris-glycine-SDS (50 mM Tris, 384 mM glycine and 0.2% SDS) as electrode buffer. SDS-PAGE was carried out at a

constant current of 16 mA per gel over 15 h. After the run, gels were stained with Sypro Ruby stain (Lonza, Rockland, ME) following the manufacturer's indications.

The 2-DE gel images were acquired with a Gel Doc XR+ system (Bio-Rad Laboratories). Image analysis of digitalized gels was performed through PDQuest Advanced software v. 8.0.1 (Bio-Rad Laboratories). 2-DE gels were matched across wild and cultivated common bean accessions. Only qualitative (presence or absence) differences among spot patterns, detectable under a sensitivity of 1.1, were analyzed.

Seed Amino Acid Composition. Protein hydrolysis was carried out on 100 mg of bean flour (average moisture content 4.32%) with 5 mL of HCl (6 N) in an ampule glass sealed for 24 h at 110 °C. After hydrolysis, the solution was diluted with 200 mL of distilled water and filtered through a 0.45 μm filter (Filter-Lab, Barcelona, Spain). Tryptophan determination was not possible, because acidic hydrolysis transformed it into ammonium. The derivatization conditions were as follows: 10 μL of sample was buffered to pH 8.8 (AccQFluor borate buffer; Waters, Milford, MA) to yield a total volume of 100 μL. Derivatization was initiated by the addition of 20 μL of AccQ-Fluor (Waters) reagent (3 mg/mL in acetonitrile).

The amino acid analyses were performed by high-performance liquid chromatography (HPLC). Chromatographic analysis conditions were as follows: separations were carried out in a Waters AccQ-Tag column (3.9 mm × 150 mm with a 4 μm of particle size); the pumps were fixed to maintain a flow rate of 1.0 mL/min; and the column heater was fixed to maintain a temperature of 37 °C. The gradient profile and composition of the mobile phase were adapted from methodology developed by Vandelen and Cohen30 and Waters AccQTag

Instruction Manual. Separation and detection were carried out using Alliance model 2695 HPLC Module (Waters) and model 2475 scanning fluorescence detector (Waters). Empower 2 Advanced software (Waters) was used to control system operation and results management. Detection was accomplished by fluorescence with excitation at 250 nm and emission at 395 nm. Amino acids were identified by retention time and quantification using an external calibration, where amino acid standard solution (AA-S-18, Sigma-Aldrich, Saint Louis, MO) was prepared to obtain a calibration curve of five points. Two biological replicate samples were carried out per accession.

Data Analysis. The degree of similarity between 2-DE phaseolin profiles of common bean populations was estimated as the percentage of spots shared through the F (also described in the literature as S) coefficient31 given by $F = 2n_{xy}/(n_x + n_y)$ where n_{xy} is the number of phaseolin spots shared by populations x and y; and where n_x and n_y are the total numbers of phaseolin spots identified in populations x and y, respectively. Dissimilarity or proteomic distance (D) between pairwise populations was then estimated as:

$$D = 1 - F$$

A hierarchical cluster analysis using the unweighted pair-group method with arithmetic averaging (UPGMA) was applied to the matrix of both pairwise proteomic distances and pairwise differences in methionine content. Nonmetric multidimensional scaling (MDS) algorithm was used to construct a two-dimensional plot on the basis of the pairwise proteomic distance matrix. The significance of the relationship between proteomic distance and difference in methionine content across pairs of accessions was assessed using Mantel's matrix comparison test, which does not require that matrix elements be

208 independent.[32] UPGMA, MDS, and the Mantel test were performed using
209 NTSYSpc v. 2.1 software (Applied Biostatistics, Inc., Setauket, NY).

210 Standard statistical analyses were performed using the statistical software
211 package IBM SPSS Statistics 20 (SPSS, Chicago, IL). Bootstrapping was used
212 to calculate 95% and 99% confidence intervals (CIs) for the means of proteomic
213 distance. For each set of N estimates of D, bootstrap samples of size N were
214 obtained following a Monte Carlo algorithm putting mass 1/N at each one of the
215 observed D estimates. From distribution of 10 000 bootstrap mean replications of
216 D, the bootstrap CIs (95% and 99%) were obtained by the bias corrected
217 percentile method.[33] A similar procedure was performed to obtain CIs for the
218 means of methionine content across accessions. Random numbers were
219 obtained using the standard multiplicative linear congruential generator
220 implemented by Schrage,[34] which uses the multiplier 16 807 and prime modulus
221 231-1 to give long sequences that have appearance of randomness. Bootstrap
222 CIs were obtained with the software DIANA (C. Zapata, unpublished).

223 ## 3. Results and discussion

224 **2-DE Phaseolin Patterns.** 2-DE phaseolin patterns in 18 wild and cultivated
225 accessions of common bean are shown in Figure 1. It can be seen that 2-DE
226 phaseolin profiles were markedly diverse across accessions. Thus, the number
227 of spots on 2-DE gels across the 18 phaseolin patterns ranged from 10 (Jalisco)
228 to 28 (Huevo de Huanchaco) and averaged (±SE, standard error) 20.4 ± 1.1.
229 Phaseolin patterns were found to be remarkably less complex in Mesoamerican
230 gene pool (M) accessions than in Andean gene pool (A) ones: the mean number
231 of spots for M and A accessions was 16.8 ± 1.4 and 23.4 ± 1.0, respectively
232 (Mann−Whitney U test, $p < 0.01$). It is noteworthy that most of these variations

were due to the presence/absence of spots in the first dimension (pI). Recent evidence has shown that differences in the pI of 2-DE phaseolin spots are largely due to differentially phosphorylated α- and β-type polypeptides.[26] The same study revealed that mobilization of the phaseolin in germinating seeds seems to occur through the degradation of highly phosphorylated isoforms. It suggests, therefore, that regulatory mechanisms responsible for the synthesis and subsequent degradation of phaseolin through phosphorylations might undergo variations over common bean populations.

Proteomic Distance. A quantitative estimation of the differences between 2-DE phaseolin profiles was obtained from the matrix of pairwise proteomic distances (D). The D-values (in percentage) between all possible pairwise comparisons of cultivated and wild common bean population are given in Figure 2. We found that proteomic distances were largely heterogeneous: values of D ranged from 0 (three M × M comparisons) to 100% (six M × A comparisons) and averaged 59.9 ± 2.4%. The mean proteomic distance (with 95% and 99% bootstrap CIs) within and between gene pools at different levels of comparison is shown in Table 2. Most of the heterogeneity in proteomic distance can be attributed to between gene pool differences. In fact, proteomic distance between gene pools was statistically significant higher than those proteomic distances within gene pools ($p < 0.01$). Within gene pools, proteomic distances were rather homogeneous over different levels of comparison with a single exception. The Mesoamerican cultivated (MC) populations exhibited indeed the lowest mean proteomic distance score (i.e., 11.0 ± 4.9), which was statistically significantly different from the remaining mean proteomic distances ($p < 0.01$), including Andean cultivated (AC) populations. Lower diversity for phaseolin types in the domesticated

Mesoamerican gene pool than in the Andean one has also been reported on the basis of 1-DE analyses.[20] This evidence would support the interpretation that a single domestication event occurred in Mexico, whereas multiple domestication events occurred in the Andes,20 but the domestication history of the common bean is still a controversial topic.[35]

On the basis of the matrix of pairwise proteomic distances, clustering analysis (UPGMA) and projection of the distances in two-dimensional space (MDS) were used to evaluate the efficiency of 2-DE phaseolin patterns in grouping the populations according to the classification by major gene pools. The UPGMA dendrogram generated from the matrix of pairwise D-values is shown in Figure 3. First, it must be highlighted that proteomic distances were able to reliably reproduce the known separation of the populations into the Mesoamerican and Andean gene pools. The only exception to the double-cluster pattern of Mesoamerican and Andean bean accessions was a well-separated branch containing the wild accession Jalisco. In addition, UPGMA dendrogram was also able to unravel heterogeneity within the two gene pools. In particular, the accession of Ayacucho formed a clearly separated subcluster within Andean cluster. The projection of the proteomic distances between pairs of the 18 common bean accessions in two-dimensional space is shown in Figure 4. The stress function value was 0.074, which indicates a good goodness-of-fit of the projected distances in the configuration space to the monotone function of the original distances. The two-dimensional map obtained by MDS reflected essentially the same scenario obtained by the dendrogram. Thus, dimension I showed noticeable differences between two tight clusters of Mesoamerican and Andean gene pools, whereas dimension II separated clearly the bean accessions

Ayacucho and Jalisco as two outliers strongly deviating from the two main clusters.

Previous investigations support the highest proteomic distance found for Ayacucho and Jalisco populations.[14,18,35−38]

First, it is important to note that Ayacucho is a nuña, a type of cultivar currently restricted to certain parts of the highlands of Peru and Bolivia that exhibits the particular property of popping when exposed to heat. It is considered that nuña beans may represent ancient preceramic landraces descendents of the earliest Andean bean domesticates.[14,36] Second, Jalisco is a possible ancestral wild accession of P. vulgaris. Recent evidence using a variety of nuclear and chloroplast DNA markers suggests that the wild forms of P. vulgaris originated in Mesoamerica and the Andean gene pool originated through different migration events from central Mexico.[18,35,37] It also seems that Mesoamerican beans were domesticated in the region of Jalisco.[38] Accordingly, analysis of genetic diversity with microsatellite markers in a large number of wild and domesticated common bean accessions from Mesoamerican and Andean gene pools has shown than the race Jalisco constituted a single separated group together with the race Durango.14 Taken together, our investigations show that proteomic distances obtained from 2-DE phaseolin patterns are a reliable marker of genetic differentiation among cultivated and wild bean populations.

Proteomic Distance and Seed Protein Quality. Essential and nonessential amino acid composition for the 18 cultivated and wild common bean seeds from the Mesoamerican and Andean gene pools is shown in Supporting Information Tables 1 and 2. Figure 5 shows that the values of the seed methionine content varied markedly across the complete set of accessions: mean values ranged from

4.2 (Contender) to 7.4 mg/g (Jalisco) of protein sample. Despite these oscillations, all accessions contain methionine-cysteine poor proteins as compared to the reference amino acid FAO/ WHO requirements recommended in human nutrition (see Supporting Information Figure 3).7 On the other hand, no statistically significant differences in methionine content were detected between Mesoamerican and Andean gene pools by the Mann−Whitney U test ($p = 0.79$). However, we found a weak relationship between the degree of proteomic distance and the difference in methionine content in total seed protein over pairs of populations (see Supporting Information Figure 4).

Specifically, a positive correlation was detected between both variables using both the Pearson product-moment correlation coefficient to assess the strength of association and the Mantel test to assess its probability ($r = 0.176$, $p < 0.05$, $n = 153$). This weak correlation can be understood taking into account that the phaseolin represents about 35−50% of the total seed protein.6 It seems, therefore, that it is difficult to predict methionine content in total seed protein from individual proteomic distance scores based on 2-DE phaseolin profiles. Cluster analysis offers a different perspective to the study of the diversity in the methionine content across bean accessions because their degree of similarity/dissimilarity can be analyzed jointly. Cluster analysis using UPGMA from differences in methionine content between pairs of accessions is shown in Figure 6. Interestingly, the accessions Jalisco and Ayacucho appear to be clear outliers distantly related to the rest of the accessions in agreement with analyses based on proteomic distances. In addition, the levels of methionine were significantly higher in seeds of Jalisco and Ayacucho than in the remaining

accessions used in this study ($p < 0.01$; Table 3), about 70% higher than in the Contender accession (Figure 5).

These observations suggest, therefore, that proteomic distances based on 2-DE phaseolin profiles are an informative biomarker to identify outliers with distinct methionine content. Accordingly, phaseolin-based proteomic distances can establish a link between P. vulgaris genetic differentiation and methionine content in total seed protein, which has important implications for the genetic improvement of bean seeds with higher nutritional quality.[5,23,39]

The study of natural variation in the protein phaseolin from 2-DE-based patterns can provide additional valuable information to improve bean cultivars. Previous studies suggest that the existing natural variation in phaseolin types could be used for improving the digestibility of common bean. Thus, it is known that the nutritional value of common bean is affected by a high resistance of phaseolin to proteolysis by the digestive enzymes.[4,5,40−43] However, the degree of hydrolysis in vitro of phaseolin seems to be influenced more by the structural characteristics of the phaseolin type determined using 1-DE than by its amino acid composition.4,5 A number of structural variations can alter the phaseolin susceptibility to proteolysis including differences in the proportion of α- and β-type polypeptides in the whole molecule and post-translational modifications such as glycosylations.[4,5,42,43] At this regard, 2-DE phaseolin profiles have the advantage that they provide more detailed information about the structural differences among phaseolin types than is available via 1-DE banding patterns. In particular, 2-DE profiles allow us to identify differentially glycosylated and phosphorylated α- and β-type phaseolin polypeptides (Supporting Information Figure 2).[25,26] It would be interesting to study as the natural variation in the levels of glycosylation and

phosphorylation of α- and β-type polypeptides affects the susceptibility of phaseolin to in vitro and in vivo proteolysis. On the other hand, it is also probable that proteomic distances based on 2-DE phaseolin patterns can provide relevant information about other bean traits. It is well established that phaseolin diversity is associated with several important agronomic traits (seed weight, tolerance to low soil pH, resistance to parasites, etc.).[11,20,21] These associations are likely a byproduct of linkage disequilibrium between alleles at phaseolin coding genes and alleles at quantitative trait loci (QTLs) of agronomic relevance. Linkage disequilibria among functionally unrelated loci could have arisen at any given moment of the history of common bean populations by evolutionary mechanisms such as genetic drift, founder effects, and stratification (or admixture) of populations, and maintained by the self-pollinated nature of P. vulgaris.[11]

4. Conclusions

Overall, our results show that proteomic distances based on high-resolution 2-DE phaseolin profiles are a reliable marker of genetic differentiation of common bean populations. Our study also indicates that phaseolin-based proteomic distances can establish a link between P. vulgaris genetic differentiation and seed protein quality as evaluated by the content of the essential amino acid methionine. This link might conceivably include other bean traits of agronomic relevance. Therefore, 2-DE phaseolin patterns might be potentially exploited to speed the process of selection of common bean germplasm in follow-up breeding programs to the improvement of bean cultivars.

Acknowledgments

This research was supported in part by grant 10PXIB262008PR (Xunta de Galicia, Spain).

The authors declare no competing financial interest.

References

1) Gepts, P.; Aragao, F. J. L.; Barros, E.; Blair, M. W.; Brondani, R.; Broughton, W.; Galasso, I.; Hernández, G.; Kami, J.; Lariguet, P.; McClean, P.; Melotto, M.; Miklas, P.; Pauls, P.; Pedrosa-Harand, A.; Porch, T.; Sánchez, F.; Sparvoli, F.; Yu, K. Genomics of Phaseolus beans, a major source of dietary protein and micronutrients in the tropics. In Genomics of Tropical Crop Plants, 1st ed.; Moore, P. H., Ming, R., Eds.; Springer Science: New York, 2008; pp 113−143.

(2) FAOSTAT. FAO statistical data base for agriculture. Food and Agriculture Organization of the United Nations (FAO). URL http:// faostat.fao.org, 2014.

(3) Krupa, U. Main nutritional and antinutritional compounds of bean seeds-A review. Polym. J. Food Nutr. Sci. 2008, 2, 149−155.

(4) Montoya, C. A.; Leterme, P.; Victoria, N. F.; Toro, O.; Souffrant, W. B.; Beebe, S.; Lallès, J.-P. Susceptibility of phaseolin to in vitro proteolysis is highly variable across common bean varieties (Phaseolus vulgaris). J. Agric. Food Chem. 2008, 56, 2183−2191.

(5) Montoya, C. A.; Lallès, J.-P.; Beebe, S.; Leterme, P. Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (Phaseolus vulgaris). Food Res. Int. 2010, 43, 443−449.

(6) Ma, Y.; Bliss, F. Seed proteins of common bean. Crop Sci. 1978, 18, 431−437.

(7) FAO/WHO/UNU. Protein and amino acid requirements in human nutrition. WHO Technical Rep. Ser. 2007, 935.

(8) Slightom, J. L.; Drong, R. F.; Klassy, C.; Hoffman, L. M. Nucleotide sequences from phaseolin cDNA clones: The major storage proteins from Phaseolus vulgaris are encoded by two unique gene families. Nucleic Acids Res. 1985, 13, 6483−6498.

(9) Pedrosa-Harand, A.; Porch, T.; Gepts, P. Standard nomenclature for common bean chromosomes and linkage groups. Annu. Rep. Bean Improv. Coop. 2008, 51, 106−107.

(10) Gepts, P. Phaseolin as an evolutionary marker. In Genetic Resources of Phaseolus Beans, 1st ed.; Gepts, P., Ed.; Kluwer Academic Publishers: Dordrech, Holland, 1988; pp 215−241.

(11) Koenig, R. L.; Singh, S. P.; Gepts, P. Novel phaseolin types in wild and cultivated common bean (Phaseolus vulgaris, Fabaceae). Econ. Bot. 1990, 44, 50−60.

(12) Singh, S. P. Broadening the genetic base of common bean cultivars: a review. Crop Sci. 2001, 41, 1659−1675.

(13) Emani, C.; Hall, T. C. Phaseolin: structure and evolution. Open Evol. J. 2008, 2, 66−74.

(14) Kwak, M.; Gepts, P. Structure of genetic diversity in the two major gene pools of common bean (Phaseolus vulgaris L., Fabaceae). Theor. Appl. Genet. 2009, 118, 979−992.

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

(16) Becerra-Velásquez, V. L.; Gepts, P. RFLP diversity in common bean (Phaseolus vulgaris L). Genome 1994, 37, 256−263.

(17) Chacón, S. M. I.; Pickersgill, B.; Debouck, D. G. Domestication patterns in common bean (Phaseolus vulgaris L.) and the origin of the Mesoamerican and Andean cultivated races. Theor. Appl. Genet. 2007, 110, 432−444.

(18) Desiderio, F.; Bitocchi, E.; Bellucci, E.; Rau, D.; Rodríguez, M.; Attene, G.; Papa, R.; Nanni, L. Chloroplast microsatellite diversity in Phaseolus vulgaris. Plant Sci. 2013, 3, 1−15.

(19) Singh, S. P.; Gepts, P.; Debouck, D. G. Races of common bean (Phaseolus vulgaris L., Fabaceae). Econ. Bot. 1991, 45, 379−396.

(20) Gepts, P.; Osborn, C.; Rashka, K.; Bliss, F. A. Phaseolin-protein variability in wild forms and landraces of the common bean (Phaseolus vulgaris): evidence for multiple centers of domestication. Econ. Bot. 1986, 40, 451−468.

(21) Johnson, W. C.; Menéndez, C.; Nodari, R.; Koinange, E. M. K.; Magnusson, S.; Singh, S. P.; Gepts, P. Association of a seed weight factor with the phaseolin seed storage protein locus across genotypes, environments, and genomes in Phaseolus-Vigna spp.: Sax (1923) revisited. J. Agric. Genomics (J. Quant. Trait Loci) 1996, 2, 5.

(22) Gepts, P.; Bliss, F. A. Enhanced available methionine concentration associated with higher phaseolin levels in common bean seeds. Theor. Appl. Genet. 1984, 69, 47−53.

(23) Kami, J. A.; Gepts, P. Phaseolin nucleotide sequence diversity in Phaseolus. I. Intraspecific diversity in Phaseolus vulgaris. Genome 1994, 37, 751−757.

(24) Salmanowicz, B. P. Phaseolin seed variability in common bean (Phaseolus vulgaris) by capillary gel electrophoresis. J. Appl. Genet. 2001, 42, 269−281.

(25) De la Fuente, M.; López-Pedrouso, M.; Alonso, J.; Santalla, M.; de Ron, A. M.; Alvarez, G.; Zapata, C. In-depth characterization of the phaseolin protein diversity of common bean (Phaseolus vulgaris L.) based on two-dimensional electrophoresis and mass spectrometry. Food Technol. Biotechnol. 2012, 50, 315−325.

(26) López-Pedrouso, M.; Alonso, J.; Zapata, C. Evidence for phosphorylation of the major seed storage protein of the common bean and its phosphorylation-dependent degradation during germination. Plant Mol. Biol. 2014, 4−5, 415−428.

(27) Biron, D. G.; Loxdale, H. D.; Ponton, F.; Moura, H.; Marché, L.; Brugidou, C.; Thomas, F. Population proteomics: An emerging discipline to study metapopulation ecology. Proteomics 2006, 6, 1712−1715.

(28) CIAT (Centro Internacional de Agricultura Tropical). Genetic Resources Program: bean collection. URL http://isa.ciat.cgiar.org/ urg/beancollection.do.

(29) De la Fuente, M.; Borrajo, A.; Bermúdez, J.; Lores, M.; Alonso, J.; Santalla, M.; de Ron, A. M.; Zapata, C.; Alvarez, G. 2-DE-based proteomic analysis of common bean (Phaseolus vulgaris L.) seeds. J. Proteomics 2011, 74, 262−267.

(30) Van Wandelen, C.; Cohen, S. Using quaternary highperformance liquid chromatography eluent systems for separating 6- minoquinolyl-N-hydroxysuccinimidyl carbamate-derivatized amino acid mixtures. J. Chromatogr., A 1997, 763, 11−22.

(31) Nei, M.; Li, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. U.S.A. 1979, 76, 5269−5273.

(32) Mantel, N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967, 27, 209−220.

(33) Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics No. 38; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1982; pp 75−92.

(34) Schrage, L. A more portable Fortran random number generator. ACM Trans. Math. Software 1979, 5, 132−138.

(35) Rossi, M.; Bitocchi, E.; Bellucci, E.; Nanni, L.; Rau, D.; Attene, G.; Papa, R. Linkage disequilibrium and population structure in wild and domesticated populations of Phaseolus vulgaris L. Evol. Appl. 2009, 2, 504−522.

(36) Tohme, J.; Toro, O. C.; Vargas, J.; Debouck, D. G. Variability in Andean nuña common beans (Phaseolus vulgaris, Fabaceae). Econ. Bot. 1995, 49, 78−95.

(37) Bitocchi, E.; Nanni, L.; Bellucci, E.; Rossi, M.; Giardini, A.; Zeuli, P. S.; Logozzo, G.; Stougaard, J.; McClean, P.; Atiene, G.; Papa, R. Mesoamerican origin of the common bean (Phaseolus vulgaris L.) is revealed by sequence data. Proc. Natl. Acad. Sci. U.S.A. 2012, 109, 5148−5149.

(38) Kwak, M.; Kami, J. A.; Gepts, P. The putative Mesoamerican domestication center of Phaseolus vulgaris is located in the Lerma-Santiago basin of Mexico. Crop Sci. 2009, 49, 554−563.

(39) Taylor, M.; Chapman, R.; Beyaert, R.; Hernández-Sebastià, C.; Marsolais, F. Seed storage protein deficiency improves sulphur amino acid content in common bean (Phaseolus vulgaris L.): redirection of sulphur from γ-glutamyl-S-methyl-cysteine. J. Agric. Food Chem. 2008, 56, 5647−5654.

(40) Marinello, L.; Giosaffato, C. V. L.; Di Pierro, P.; Sorrentino, A.; Porta, R. Synthesis and resistance to in vitro proteolysis of transglutaminase cross-linked phaseolin, the major storage protein from Phaseolus vulgaris. J. Agric. Food Chem. 2007, 55, 4717−4721.

(41) Marquez, U. M.; Lajolo, F. M. Composition and digestibility of albumin, globulins and glutelins from Phaseolus vulgaris. J. Agric. Food Chem. 1981, 29, 1068−1074.

507 (42) Genovese, M. I.; Lajolo, F. M. Influence of naturally acid-soluble proteins
508 from beans (Phaseolus vulgaris L.) on in vitro digestibility determination. Food
509 Chem. 1998, 62, 315−323.
510 (43) Jivotovskaya, A. V.; Senyuk, V. I.; Rotari, V. I.; Horstmann, C.; Vaintraub, I.
511 A. Proteolysis of phaseolin in relation to its structure. J. Agric. Food Chem. 1996,
512 44, 3768−3772.
513

CAPTIONS TO FIGURE

Figure 1. High-resolution 2-DE phaseolin profiles from seeds of 18 cultivated and wild accessions of common bean corresponding to the Mesomerican and Andean gene pools. Positions of the spots according to their isoelectric point (pI) and molecular mass (Mr) are shown.

Figure 2. Pairwise proteomic distance (D), expressed as a percentage, based on high-resolution 2-DE phaseolin patterns from seeds of 18 cultivated and wild accessions of common bean.

Figure 3. Unweighted pair-group method with arithmetic averaging (UPGMA) dendrogram from the matrix of pairwise proteomic distances (D) based on high-resolution 2-DE phaseolin profiles from seeds of 18 cultivated and wild accessions of common bean

Figure 4. Multidimensional scaling (MDS) projection of the proteomic distance (D) between 18 cultivated and wild populations onto two-dimensional space based on high-resolution 2-DE phaseolin patterns.

Figure 5. Mean methionine content (mg/g total seed protein) across 18 cultivated and wild populations of common bean.

Figure 6. Unweighted pair-group method with arithmetic averaging (UPGMA) dendrogram based on pairwise differences in seed methionine content (mg/g total protein) from 18 cultivated and wild populations of common bean.

535 **Table 1.** Information on the Different Cultivated and Wild Bean Accessions Used

536 in This Study

accession identifier	common name	phaseolin[b]	gene pool[c]	race[d]	status	origin/ country	origin/ locality	seed/weight of 100 seeds (g)	seed/color
PI 201011 (G 12851)		S, Sb, Sd, M11	M		wild	Guatemala	Santa Rosa	6.0	cream, black, other
PI 268110 (G 1459)	Jamapa	S	M	M	landrace	Mexico	Veracruz	18.0	black
PI 313590 (G 2511)	Boyaca 22	B, S	M	M	cultivated	Colombia	Boyaca	16.0	brown, black
PI 417780 (G 12953)		M	M		wild	Mexico	Jalisco	10.5	brown
PI 549695 (G 4498)	Sanilac	S	M	M	cultivated	U.S.		14.0	white
PI 661723			M		wild	Costa Rica	Puntarenas		brown
PI 661840 (G 22837)		M4, Sb	M		wild	Mexico	Chihuahua	11.1	cream, brown
W6 35834 (G 5773)	Ica Pijao	B	M	M	cultivated	Colombia	Valle Del Cauca	24.0	black
G 12053	Nuña	A	A	P	cultivated	Perú	Ayacucho	36.6	purple
G 12588	Huevo de Huanchaco (nuña)	H	A	P	cultivated	Perú	Cajamarca	49.8	white, red
PI 390770 (G 7225)		T	A		wild	Perú	Apurimac	11.0	cream, black
PI 474218 (G 6412)	Contender	C	A	NG	cultivated	U.S.		44.0	cream, brown
PI 549633 (G 7476)	Tendergreen	T	A	NG	cultivated	U.S.		30.0	cream, brown
PI 557465 (G 4474)	Coscorron	S, B	A	C	landrace	Chile		52.0	white, other
PI 638874			A		wild	Argentina	Jujuy		brown, black
PI 640968			A		wild	Argentina	Salta		brown, black
PI 661817			A		wild	Bolivia	Chuquisaca		cream
W6 35784 (G 4494)	Calima	T	A	NG	cultivated	Colombia		54.8	red, cream

537

538

Table 2. Proteomic Distance (D) within and between Common Bean Gene Pools Based on 2-DE Phaseolin Profiles aM, Mesoamerican gene pool populations; A, Andean gene pool populations; MC, Mesoamerican cultivated populations; MW, Mesoamerican wild populations; AC, Andean cultivated populations; AW, Andean wild populations. bNumber of accession pairs. cData (expressed in percentage) are means ± standard error. dCI, confidence interval; CL, lower bound; CU, upper bound.

comparison level[a]	n^b	D^c	95% bootstrap CI (CL, CU)[d]	99% bootstrap CI (CL, CU)
		Within Gene Pools		
M × M	28	37.8 ± 5.6	27.1, 49.4	23.4, 52.7
A × A	45	34.5 ± 2.9	29.2, 40.0	27.8, 41.5
MC × MC	6	11.0 ± 4.9	3.7, 18.3	0.0, 22.0
MW × MW	6	56.3 ± 13.1	34.0, 77.2	25.8, 86.0
MC × MW	16	40.8 ± 6.9	28.4, 54.9	25.6, 59.5
AC × AC	15	38.1 ± 6.2	26.3, 50.0	23.1, 52.5
AW × AW	6	30.2 ± 1.4	27.5, 32.5	26.8, 33.0
AC × AW	24	33.0 ± 3.8	26.4, 40.6	24.8, 42.7
		Between Gene Pools		
M × A	80	81.8 ± 1.1	79.6, 84.1	78.7, 84.7
MC × AC	24	79.5 ± 1.8	76.2, 83.2	74.7, 83.9
MC × AW	16	75.2 ± 3.1	69.3, 81.2	67.6, 83.1
MW × AC	24	86.0 ± 1.4	83.3, 88.8	82.6, 89.5
MW × AW	16	85.6 ± 2.4	81.3, 90.1	79.5, 91.7

Table 3. Methionine Content (mg/g protein) in Total Seed Protein of Ayacucho and Jalisco and Other Accessions

accessions	n^a	methionine content[b]	95% bootstrap CI (CL, CU)[c]	99% bootstrap CI (CL, CU)
Ayacucho	2	6.95 ± 0.15	6.94, 7.10	6.80, 7.10
Jalisco	2	7.44 ± 0.45	7.43, 7.88	6.99, 7.88
other accessions	32	5.22 ± 0.13	4.98, 5.47	4.91, 5.52

[a]Number of values. [b]Data are means ± standard error. [c]CI, confidence interval; CL, lower bound; CU, upper bound.

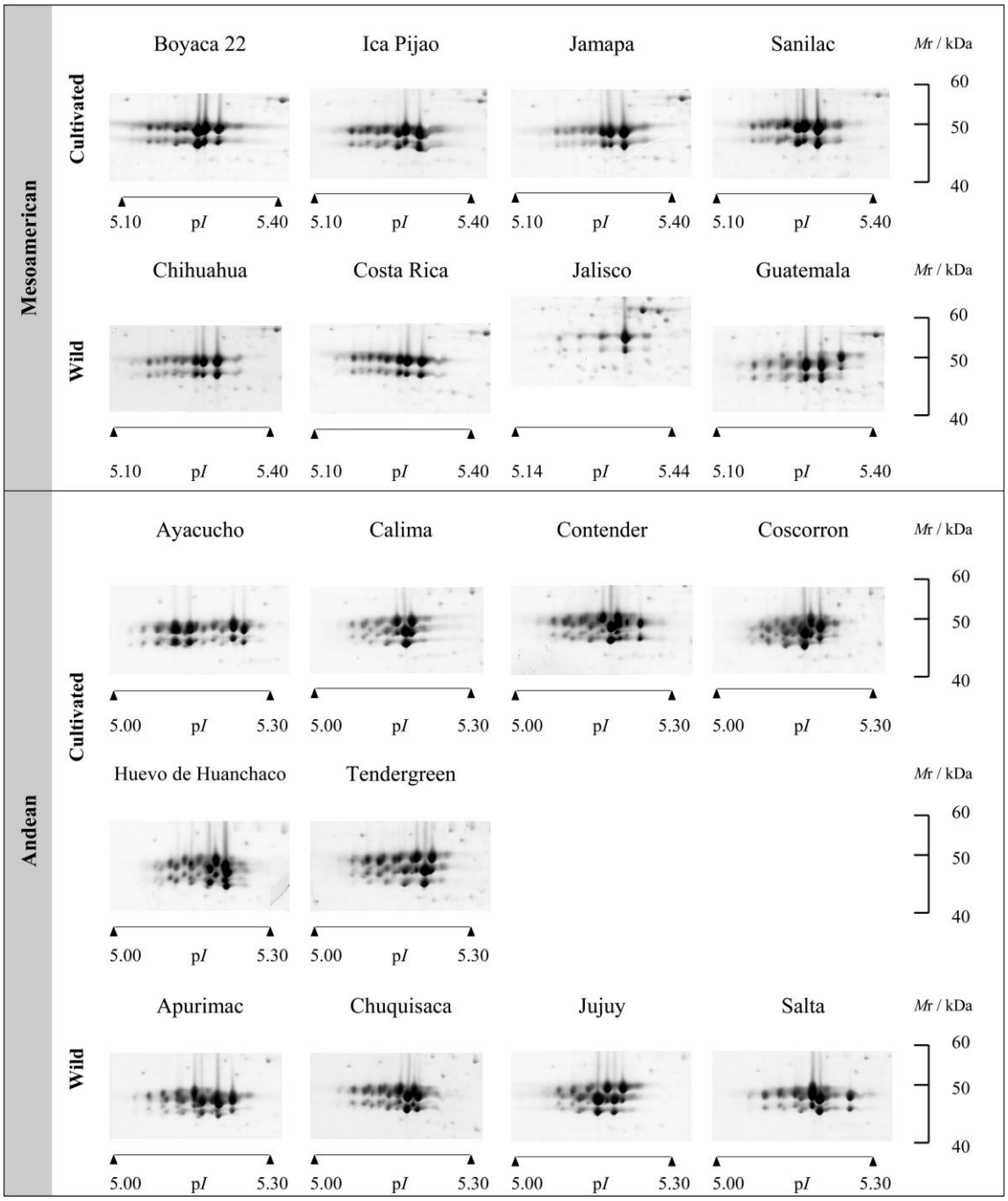

Mesoamerican
Andean
Cultivated
Wild
Boyaca 22
Ica Pijao
Jamapa
Sanilac
Chihuahua
Costa Rica
Jalisco
Guatemala
Ayacucho
Calima
Contender
Coscorron
Huevo de Huanchaco
Tendergreen
Apurimac
Chuquisaca
Jujuy
Salta
Mr / kDa
60
50
40
pI
5.10
5.40
5.14
5.44
5.00
5.30

Figure 1.

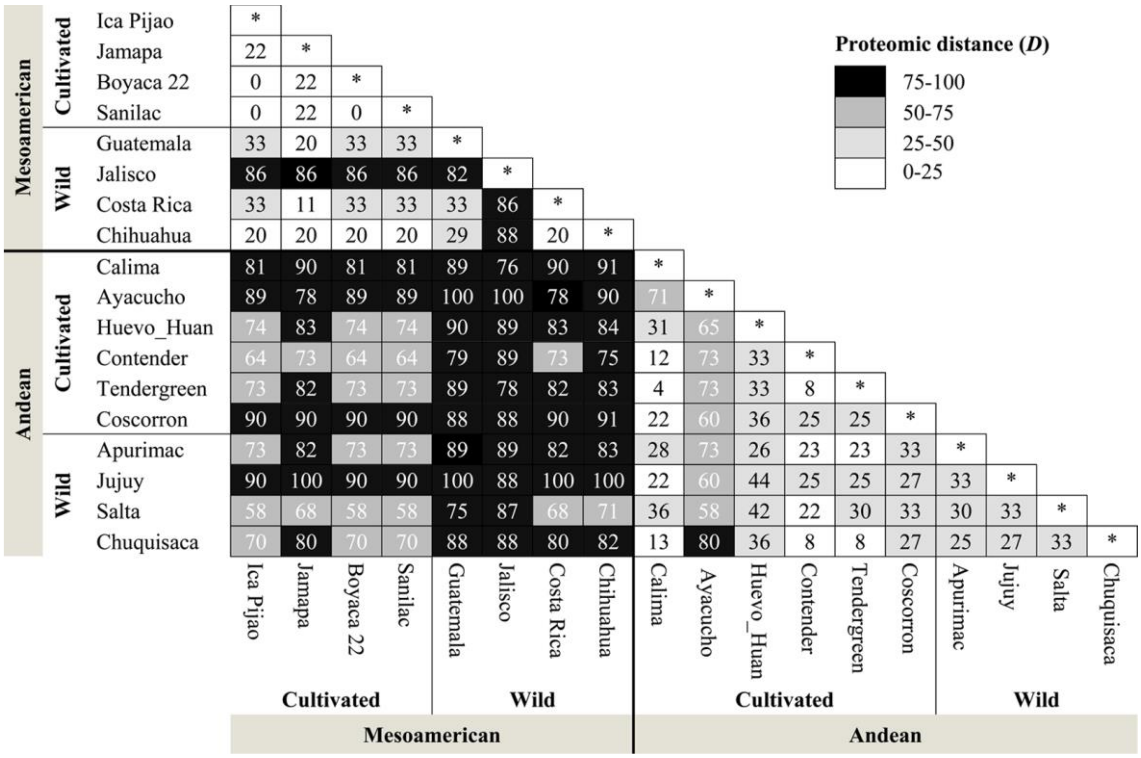

Proteomic distance (D)
75-100
50-75
25-50
0-25
Mesoamerican
Andean
Cultivated
Wild
Ica Pijao
Jamapa
Boyaca 22
Sanilac
Guatemala
Jalisco
Costa Rica
Chihuahua
Calima
Ayacucho
Huevo_Huan
Contender
Tendergreen
Coscorron
Apurimac
Jujuy
Salta
Chuquisaca

Figure 2.

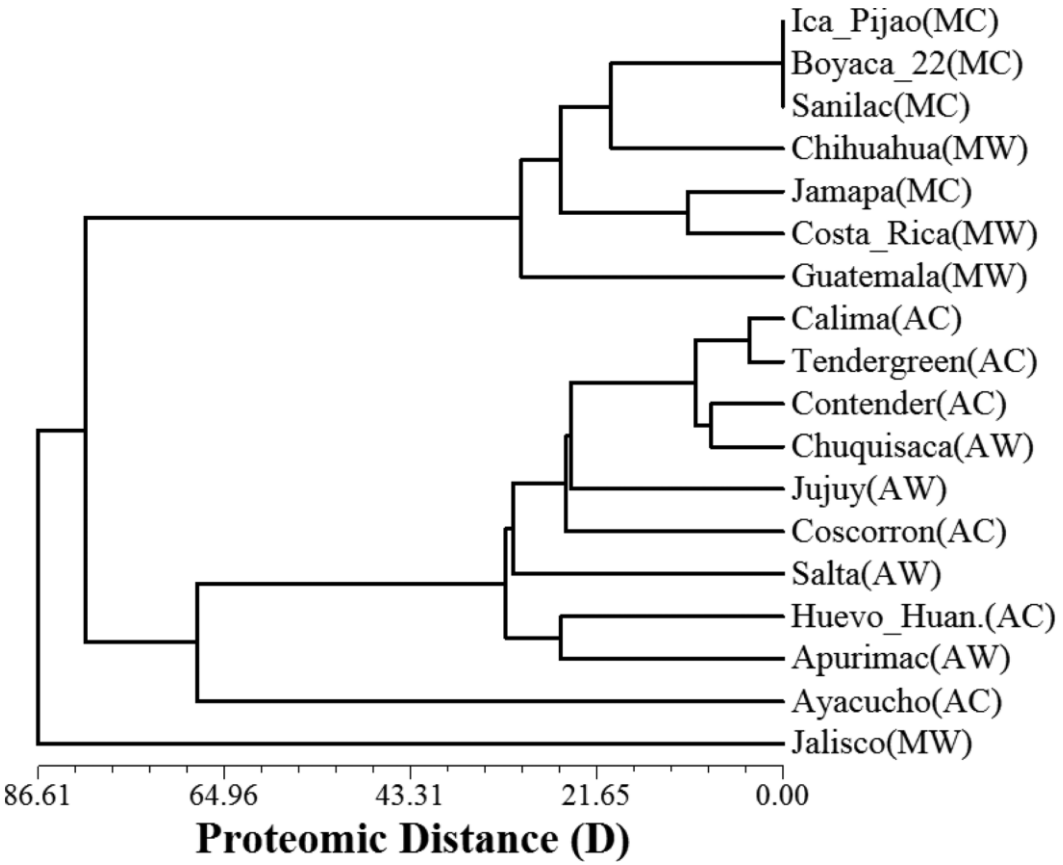
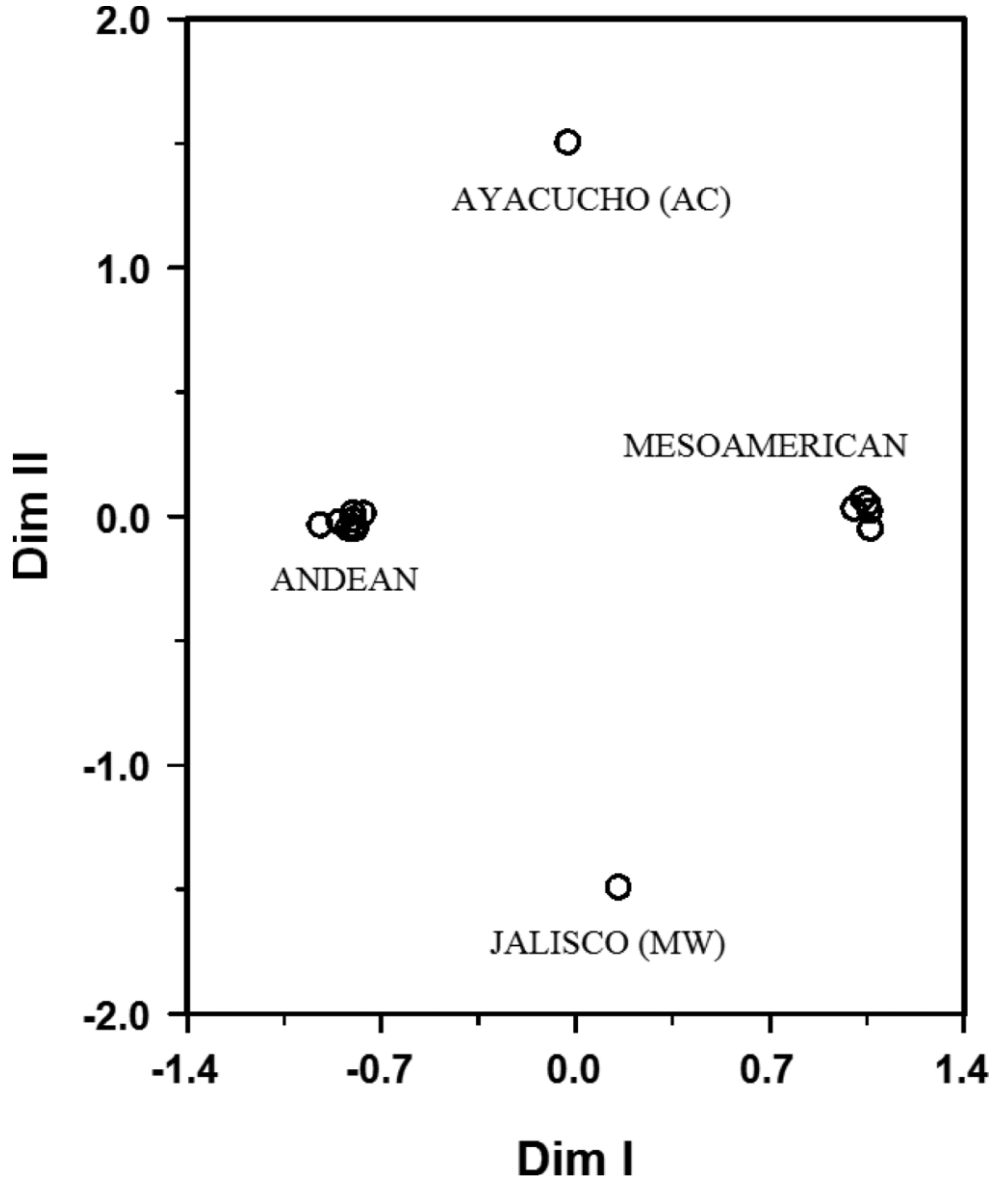

Ica_Pijao(MC)
Boyaca_22(MC)
Sanilac(MC)
Chihuahua(MW)
Jamapa(MC)
Costa_Rica(MW)
Guatemala(MW)
Calima(AC)
Tendergreen(AC)
Contender(AC)
Chuquisaca(AW)
Jujuy(AW)
Coscorron(AC)
Salta(AW)
Huevo_Huan.(AC)
Apurimac(AW)
Ayacucho(AC)
Jalisco(MW)
86.61
64.96
43.31
21.65
0.00
Proteomic Distance (D)

Figure 3.

604



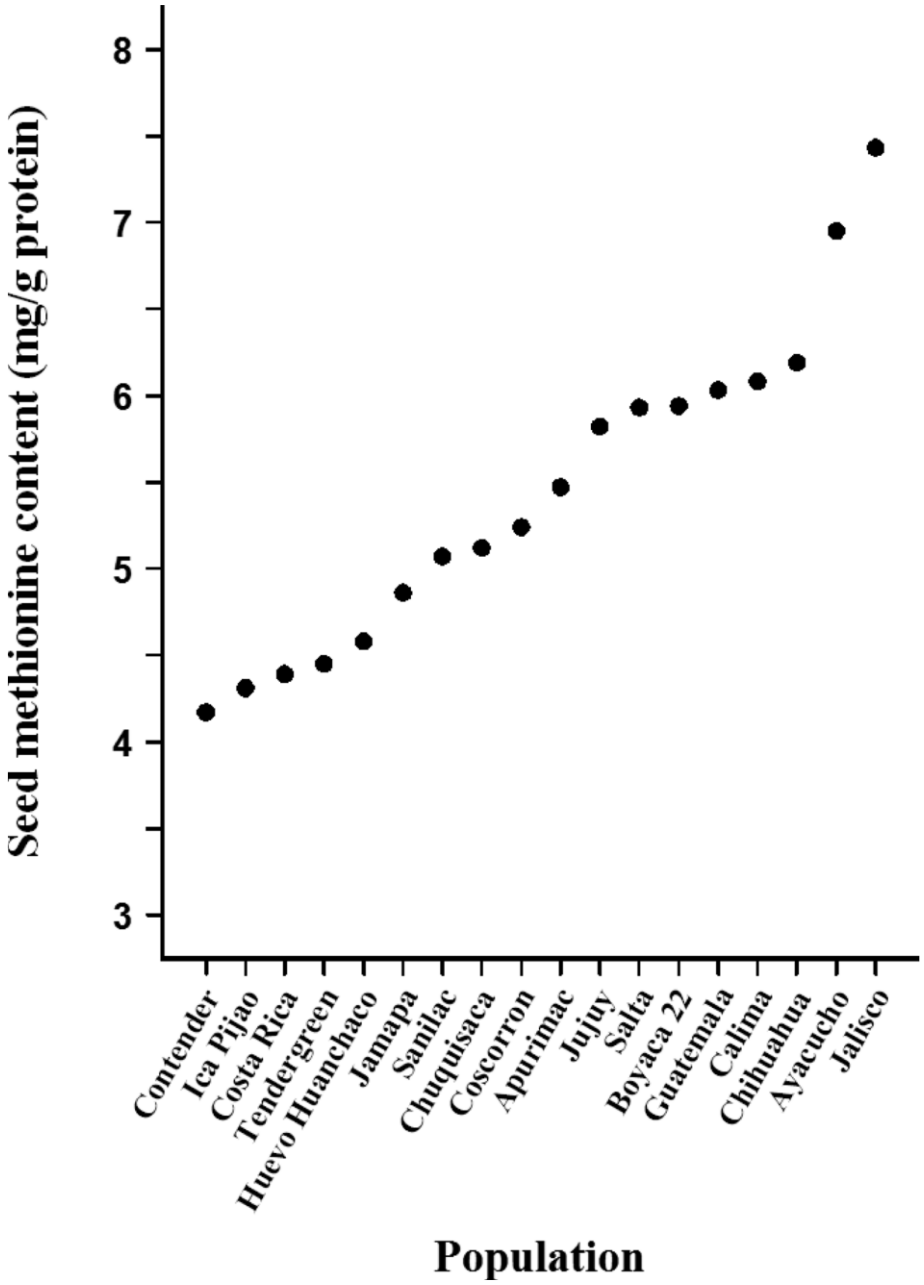
605

606 **Figure 4.**

607

608

609

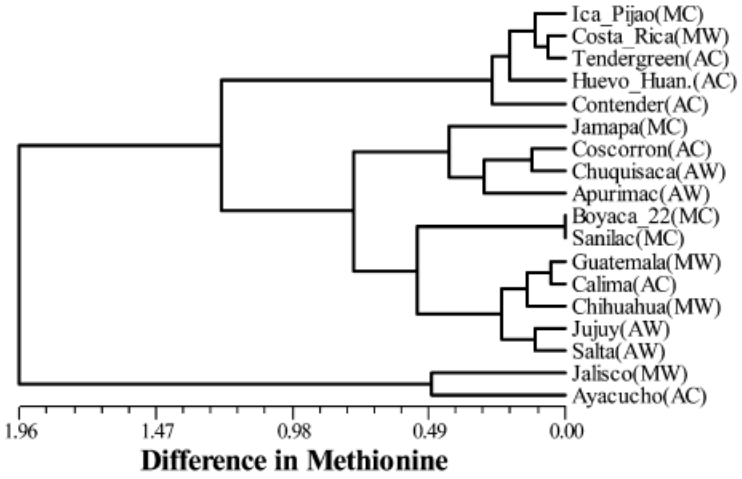


610

611 **Figure 5.**

612

613



614

615 **Figure 6.**

616

617