http://journal.unoeste.br/index.php/ca/index
DOI: 10.5747/ca.2022.v18.n2.a484
ISSN on-line 1809-8215

Colloquium Agrariae

Submetido: 11/02/2022 Revisado: 07/03/2022 Aceito: 22/04/2022

Core collection of common bean formed from traditional Brazilian germplasm

Paula Pereira Torga[1], Flávio Breseghello[1], Erica Munique da Silva[2], Rosana Pereira Vianello[1], Maria José Del Peloso[1], Leonardo Cunha Melo[1], Joaquim Geraldo Caprio da Costa[1], Silvando Carlos da Silva[1], Helton Santos Pereira[1], Aluana Gonçalves de Abreu[1]

[1]Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA Arroz e Feijão, Santo Antônio de Goiás, GO. [2]Universidade Federal de Goiás – UFG, GO. E-mail: paula.torga@embrapa.br

Abstract

The aim of this study was to select traditional accessions, compose a core collection of common bean, and assess the representativeness of the collection in relation to the base collection accommodated in the BAG of Embrapa using analysis strategies for multivariate models. We used data characterizing 2903 accessions from collections representing all geographic areas of Brazil regarding three morphologic descriptors (seed color, growth habit type, and seed size) and four ecogeographic descriptors (geographical areas, states, altitudes, and soil classes). A set of 400 accessions were selected using multivariate models applied to the data transformed in multibinary values. The accessions sampled had maximum similarity (100%) to the traditional collection, phenotypic diversity, and representative heterogeneity in relation to the traditional collection. In the core collection, the accessions represented 9.5% of the traditional accessions and were equivalent to 3% of the accessions of the base collection. Thus, it is possible to form a core collection that is representative of the base collection regarding genetic diversity and the conservation of rare alleles.

Keywords: similarity coefficient; phenotypic diversity; heterogeneity index; multivariate models; *Phaseolus vulgaris* L.

Coleção nuclear de feijão comum formada por germoplasma tradicional de coletas brasileiras

Resumo

O objetivo deste trabalho foi selecionar acessos tradicionais, compor uma coleção nuclear de feijoeiro comum e avaliar sua representatividade em relação à coleção base de coletas hospedadas no BAG da Embrapa. Utilizando dados de caracterização de 2903 acessos de coletas representando todas as regiões geográficas do Brasil quanto a três descritores morfológicos (cor de semente, tipos de crescimento e tamanho de semente) e quatro descritores ecogeográficos (regiões geográficas, unidades federativas, altitudes e classes de solos), foram selecionados 400 acessos utilizando modelos multivariados aplicados aos dados transformados em valores multibinários. Os acessos amostrados apresentaram similaridade máxima (100%) com a coleção tradicional, diversidade fenotípica e heterogeneidade representativa em relação à coleção tradicional. Na coleção nuclear, os acessos representaram 9,5% dos acessos tradicionais e equivalem a 3% dos acessos da coleção base. Com isso conclui-se que é possível formar uma coleção nuclear representativa da coleção base, no que diz respeito à diversidade genética e a conservação de alelos raros.

Palavras-chave: coeficiente de similaridade; diversidade fenotípica; índice de heterogeneidade; modelos multivariados; *Phaseolus vulgaris* L.

Introduction

Information regarding the conservation, variability, and genetic divergence within a species are essential for rational use of genetic resources (NASS, 2007). With the concern of preserving genetic diversity, many countries have collected and stored their germplasm, resulting in a large number of collections worldwide. Some collections are so large that conservation, evaluation, and accessibility to genetic diversity become difficult. Frankel and Brown (1984) proposed the concept of a "core collection" that aims at ensuring all the genetic variability of the base collection in a reduced number of accessions. Thus, there is greater facility in effectively using the genetic resources (NAYAK *et al*., 2014; MARCHINI, 2019), and that encourages researchers to more frequently draw on the stored germplasm. Consequently, various authors have presented strategies and sampling methods in selection of accessions to obtain a core collection (CHANDRA *et al*., 2002; LI *et al*., 2004; ZEWDIE *et al*., 2004; ABADIE *et al*., 2005; VASCONCELOS *et al*., 2007; VASCONCELOS *et al*., 2010). These principles and methodologies concerning the ideal size and the appropriate choice of individuals are still controversial. Various studies have suggested proportions ranging from 5% to 30% (ZHANG *et al*., 2010; ODONG *et al*., 2013; LIU *et al*., 2015). Other authors have suggested minimum samplings of 2% or less (JIANG *et al*., 2010; UPADHYAYA *et al*., 2010; SHARMA *et al*., 2012; WANG *et al*., 2014), as long as suitable mathematical models are used.

Among the various methods of selection of individual accessions to form a core collection, stratified sampling has been most used, due to the lack of or little information regarding characterization of the accessions (ABADIE *et al*, 1999, 2005). Such a method consists of three steps: (1) the accessions may be allocated in each group (stratum) taking origin, passport data, and phenotypic, genotypic, or ecogeographic data into consideration; (2) the number of accessions is allocated in each group using the direct or logarithmic proportion or combinations between the direct proportion and an index (e.g., direct proportion plus the Shannon index); and (3) the individuals of each group are chosen at random or through genetic information among accessions (FRANCO *et al*., 2005; SILVA, 2016). In addition, several methods and software programs have been cited in the literature that can assist in the formation of a core collection (GOUESNARD *et al*., 2001; FRANCO *et al*., 2006; VASCONCELOS *et al*., 2007; KIM *et al*., 2007; POLICRITI; SGARRO, 2011).

The establishment of core collections has generally taken into consideration only the information available from passport, phenotypic, and/or ecogeographic data. In this case, the collections have been validated after their formation by genetic analyses (SILVA *et al*., 2019). However, it is known that in any collection, traditional accessions important for the local economy must be present. This action allows the use and study of germplasm selected by traditional knowledge, representing the strength of local agriculture (RANGEL *et al*., 2013; BORSUK *et al*., 2017).

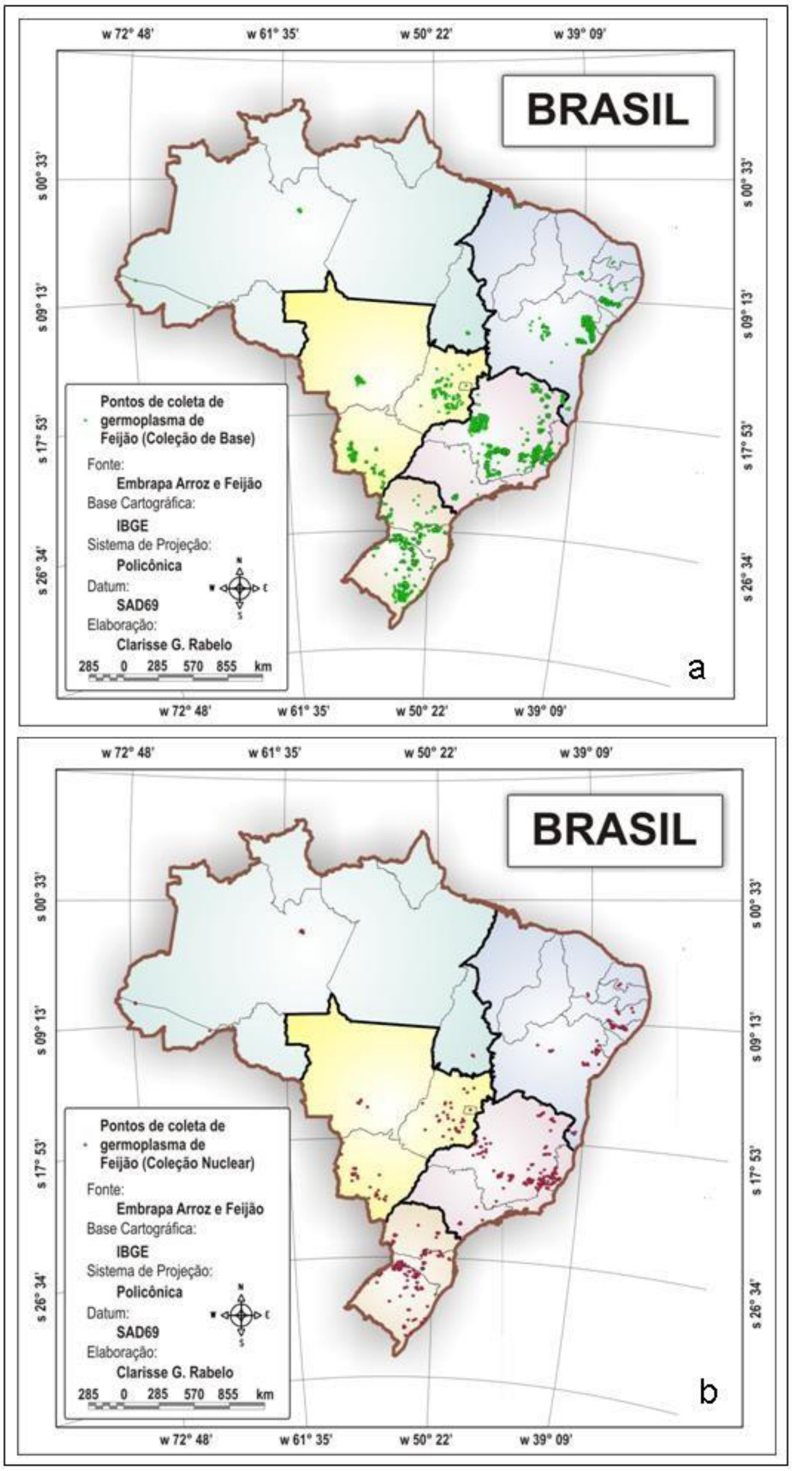
The genetic variability that exists in the germplasm of traditional accessions of common bean (*Phaseolus vulgaris* L.) in the Active Germplasm Bank (*Banco Ativo de Germoplasma* - BAG) of Embrapa Arroz e Feijão has been recognized and analyzed in various studies (BURLE *et al*., 2010; KNUPP *et al*., 2017; SILVA *et al*., 2017; VALDISSER *et al*., 2017; GRIGOLO *et al*., 2018). This variability is of fundamental importance for common bean breeding programs. They require continual acquisition of desirable alleles present in the genetic resources. The aim of this study was to select traditional accessions, compose a core collection of common bean, and evaluate the representativeness of this collection in relation to the base collection accommodated in the BAG of Embrapa.

Materials and Methods

In the current *P. vulgaris* L. germplasm collection of Embrapa Arroz e Feijão, composed of 14,307 accessions, 4,547 are traditional accessions originating from collections made in Brazil (Figure 1a). The high costs and difficult handling of all these accessions have led the bean breeding team to propose the formation of a core collection that would represent the genetic variability of the base collection of traditional accessions or collections. The size of the core collection was established, taking into consideration its representativeness of the common bean germplasm and the ease of use of the genetic variability contained in the base collection, as well as ease of use concerning management practices at field level, above all, facilitating morphological and agronomic characterization. The core collection is composed of 400 traditional accessions, with adaptation to local geoclimatic conditions (Figure 1b).

In relation to the 400 accessions that represented the traditional accessions, the data from characterization of 2903 accessions of collections representing all geographic regions of Brazil were used (Figure 1a) regarding three morphological descriptors: seed color (twelve colors), growth habit type (six types), and seed size (four intervals of 100-seed weight ranging from 10.01 g to 61.73 g). Four ecogeographic descriptors were also used: geographic regions (five regions), states (21 units), altitudes (11 altitudes ranging from 3 m to 1203 m, with divisions every 100 m), and soil classes (27 classes). These data were worked on jointly by multivariate models.

Figure 1. Map of Brazil with collection sites of 4,547 traditional common bean (*P. vulgaris* L.) accessions **(a)**, and collection sites of 400 traditional accessions selected by the multivariate analysis strategy **(b)**.



BRAZIL; Sites of collection of common bean germplasm (Base Collection / Core Collection); Source: Embrapa Arroz e Feijão; Cartographic base: IBGE; Projection System: Polyconic; Datum: SAD69; Prepared by: Clarisse G. Rabelo

For multivariate analysis, the phenotypic and ecogeographic data were transformed into multibinary data, and the similarity coefficient for qualitative variables was used. The coefficients for these variables are normally concentrated on measuring similarity, based on counting the agreements (positive or negative) between the elements. According to Bussab *et al*. (1990), when a qualitative variable has more than two levels, it is possible to transform it into binary variables through the creation of fictitious variables. Let *y´* be a qualitative variable formed by a vector of nominal qualitative variables: $y' = (y_1, y_2, ..., y_l)$, where the *i*-th component assumes l_i levels, codified such that $y_i = j$, with $j = 1, 2, ..., l_i$. Upon transforming this variable into a binary variable, each component (qualitative criterion) will contribute to generation of l_i binary variables $x_k(i)$ such that $x_i(i) = 1$ if $y_i = k$ or $x_i(i) = 0$ if $y_i \neq k$. Thus, the vector *y* of dimension *n* is transformed into vector *x* of dimension *p*, formed by binary components and, consequently, *y´* is transformed into *x´*:

$$y' = (y_1, y_2, ..., y_l) \rightarrow x' = (\underbrace{0,...,\overset{y_1}{1},...,0}_{l_1};...;\underbrace{0,...,\overset{y_l}{1},...,0}_{l_i})$$

. Such vectors of binary variables allow generation of a double entry table, making it possible to estimate the similarity coefficient between two accessions. In this case, from the values generated by the double entry table, the similarity coefficient was estimated using the Jaccard coefficient, which measures positive agreements between the elements, with similarity in the interval $[0,1]$. Such values were used in cluster analysis after transformation of the similarity coefficients into dissimilarity coefficients. The centroid method (BUSSAB *et al*., 1990) was used for cluster analysis, generating 172 groups. The number of accessions selected within each group was proportional to its size, while in the strata, with a reduced number of accessions, at least one accession was sampled. All the statistical calculations were implemented in the SAS (Statistical Analysis System) computational system through the *proc distance* and *proc cluster* procedures (SAS INSTITUTE, 2002).

To evaluate the quality of the collection formed by traditional accessions, three statistics of an exploratory, that is, non-inferential, nature were used: the similarity coefficient (CR%) proposed by Hu *et al*. (2000), with an adaptation for data of a qualitative nature, given by $CR\% = 1/m\left(\sum_{j=1}^{m} C_j^N / C_j^B\right) x100$, where m is the number of variables or descriptors $(j = 1, \; 2, \ldots, m)$ and C_j is the number of categories associated with variables j, manifested in the core collection $\left(C_j^N\right)$ and in the base collection $\left(C_j^B\right)$; the phenotypic diversity coefficient (PD) (SOUZA; PEÑALOZA, 2005), given by $DF = \sum_{j=1}^{m} D_j^N \Big/ \sum_{j=1}^{m} D_j^B$, in which $D_j = 1 - \sum_{k=1}^{C_j} f_k^2 \Big/ 1 - \frac{1}{C_j}$, where D_j is the diversity index associated with variable j, manifested in the core collection $\left(D_j^N\right)$ and in the base collection $\left(D_j^B\right)$, with $C_j \geq 2$ categories and relative frequencies $f_k(k = 1, 2, \ldots, C_J \; categorias)$ – the index D_j has variation between zero (all the observations or accessions in the same category of the variable) and one (observations equally distributed in the C_j categories of the variable), being accumulated for the “m” variables in the core collection (CC) and in the base collection (BC) to generate the PD coefficient, with variation between zero and $p \Big/ \sum_j D_j^B$; and heterogeneity, measured by the Shannon index (H$'$) (SHANNON; WEAVER, 1949), given by $H' = -\sum_{e=1}^{S} p_e \ln p_e; \;\; p_e = n_e / N$, where p_e is the relative abundance of species *e* (or group), n_e is the number of individuals of the species *e* (or group), N is the total number of individuals, and S is the total number of species (or groups).

Results and Discussion

The multivariate model analysis strategy provided an adequate sampling of the number of traditional accessions, both in the phenotypic descriptors (Table 1) and in the ecogeographic descriptors (Table 2, 3, and 4) and ensured the representativeness of accessions in all the strata.

These tables showed that this type of sampling can increase the probability of capturing the less frequent alleles, avoiding the excessive sampling of accessions of the strata with high frequency, as well as provide good sampling of the low frequency groups. In this case, in the grain color group stratum, 72% of the white accessions were sampled (Table 1). In the base collection, the frequency of accessions of growth habit type II was more than 50%, whereas only 8% were sampled to compose the core collection. The same occurred for the small seed stratum, where likewise only 8% of the accessions were sampled (Table 1).

Table 1. Distribution of absolute frequency (AF) and relative frequency (RF%) of the collection of traditional accessions of common bean (*P. vulgaris* L.) of BAG – Embrapa Arroz e Feijão and of the sampling of traditional accessions according to the strata formed regarding phenotypic variables.

	Collection of Traditional Accessions		**Sampling of Traditional Accessions**	
Stratum	**AF**	**RF%**	**AF**	**RF%**
		Grain color group		
White	18	0.62	13	3.25
Biege (*Mulatinho*)	745	25.66	62	15.50
Yellow	166	5.72	38	9.50
Brown	267	9.20	37	9.25
Pink	81	2.79	23	5.75
Purple	257	8.85	30	7.50
Red	65	2.24	15	3.75
Black	803	27.66	50	12.50
Carioca	126	4.34	28	7.00
Pinto	182	6.27	49	12.25
OTHERS	151	5.20	43	10.75
Jalo	42	1.45	12	3.00
Total	**2903**	**100**	**400**	**100**
		Growth habit type		
Determinate, type I	473	16.29	94	23.50
Indeterminate, type II	1535	52.88	123	30.75
Indeterminate, type III	513	17.67	87	21.75
Indeterminate, type IV	67	2.31	15	3.75
Indeterminate between II and III	303	10.44	71	17.75
Indeterminate between III and IV	12	0.41	10	2.50
Total	**2903**	**100**	**400**	**100**
		Seed size		
Very large (> 40.0 g)	91	3.13	40	10.00
Large (30.1 to 40.0 g)	326	11.23	95	23.75
Medium (21.0 to 30.0 g)	1052	36.24	151	37.75
Small (< 21.0 g)	1434	49.40	114	28.50
Total	**2903**	**100**	**400**	**100**

In the ecogeographic variables, the strata with only one accession were fully sampled (Table 2). In the soil classes, the greatest frequency of the accessions occurred in *Cambissolo Háplico* soil and few accessions occurred at altitudes above 1000 m (Tables 3 and 4). Therefore, this type of sampling may lead to a reduction in the biases brought about by the size of the group. In this case, the low-frequency accessions increased and the high-frequency decreased by approximately 50% in relation to the frequencies contained in the traditional collection, with all being sampled. Furthermore, sampling by the multivariate model can contribute to an increase in the diversity of the core collection.

Table 2. Distribution of the absolute frequency (AF) and relative frequency (RF%) of the collection of traditional accessions of common bean (*P. vulgaris* L.) of BAG – Embrapa Arroz e Feijão and of the sampling of traditional accessions according to the strata formed regarding the ecogeographic variables of geographic region and Brazilian state.

		Base Collection		**Core Collection**	
State	**Abbrev.**	**AF**	**RF%**	**AF**	**RF%**
		West Central			
Distrito Federal	DF	1	0.03	1	0.25
Goiás	GO	93	3.20	28	7.00
Mato Grosso do Sul	MS	142	4.89	26	6.50
Mato Grosso	MT	52	1.79	5	1.25
Total		288	9.91	60	15.00
		Northeast			
Alagoas	AL	147	5.06	21	5.25
Bahia	BA	226	7.79	26	6.50
Ceará	CE	5	0.17	2	0.50
Pernambuco	PE	5	0.17	5	1.25
Rio Grande do Norte	RN	2	0.07	2	0.50
Sergipe	SE	1	0.03	1	0.25
Total		386	13.29	57	14.25
		North			
Acre	AC	4	0.14	4	1.00
Amazonas	AM	5	0.17	4	1.00
Maranhão	MA	3	0.10	1	0.25
Rondônia	RO	1	0.03	1	0.25
Tocantins	TO	6	0.21	2	0.50
Total		19	0.65	12	3.00
		Southeast			
Espirito Santo	ES	123	4.24	35	8.75
Minas Gerais	MG	1401	48.26	94	23.50
São Paulo	SP	15	0.52	7	1.75
Total		1539	53.02	136	34.00
		South			
Paraná	PR	74	2.55	24	6.00
Rio Grande do Sul	RG	422	14.54	82	20.50
Santa Catarina	SC	175	6.03	29	7.25
Total		671	23.12	135	33.75
Total		**2903**	**100**	**400**	**100**

Table 3. Distribution of the absolute frequency (AF) and relative frequency (RF%) of the collection of traditional accessions of common bean (*P. vulgaris* L.) of BAG – Embrapa Arroz e Feijão and of the sampling of traditional accessions according to the strata formed regarding the ecogeographic variable of soil classes.

	Base Collection		**Core Collection**	
Stratum	**AF**	**RF%**	**AF**	**RF%**
Rock outcrops	9	0.31	3	0.75
Alissolo Crômico	1	0.03	1	0.25
Argilossolo Acinzentado	4	0.14	1	0.25
Argilossolo Amarelo	7	0.24	4	1.00
Argilossolo Vermelho	353	12.16	49	12.25
Argilossolo Vermelho-Amarelo	672	23.15	59	14.75
Cambissolo Háplico	731	25.18	102	25.50
Cambissolo Húmico	5	0.17	3	0.75
Chernossolo Argilúvico	24	0.83	5	1.25
Chernossolo Rêndzico	24	0.83	6	1.50
Sand dunes	1	0.03	1	0.25
Espodossolo Ferrocárbico	1	0.03	1	0.25
Gleissolo Háplico	3	0.10	1	0.25
Latossolo Amarelo	63	2.17	13	3.25
Latossolo Bruno	35	1.21	11	2.75
Latossolo Vermelho	211	7.27	27	6.75
Latossolo Vermelho-Amarelo	505	17.40	58	14.50
Luvissolo Crômico	4	0.14	2	0.50
Neossolo Flúvico	1	0.03	1	0.25
Neossolo Litólico	57	1.96	16	4.00
Neossolo Quartzarênico	55	1.89	5	1.25
Neossolo Regolítico	19	0.65	3	0.75
Nitossolo Háplico	5	0.17	2	0.50
Nitossolo Vermelho	51	1.76	13	3.25
Planossolo Háplico	13	0.45	4	1.00
Planossolo Hidromórfico	31	1.07	6	1.50
Vertissolo Ebânico	15	0.52	3	0.75
Without classification	3	0.10	-	-
Total	**2903**	**100**	**400**	**100**

Table 4. Distribution of the absolute frequency (AF) and relative frequency (RF%) of the collection of traditional accessions of common bean (*P. vulgaris* L.) of BAG – Embrapa Arroz e Feijão and of the sampling of traditional accessions according to the strata formed regarding the ecogeographic variable of altitude.

	Base Collection		**Core Collection**	
Stratum	**AF**	**RF%**	**AF**	**RF%**
0 ----| 100	142	4.89	35	8.75
100 ----| 200	315	10.85	48	12.00
200 ----| 300	184	6.34	33	8.25
300 ----| 400	366	12.61	51	12.75
400 ----| 500	122	4.20	18	4.50
500 ----| 600	203	6.99	46	11.50
600 ----| 700	340	11.71	54	13.50
700 ----| 800	316	10.89	54	13.50
800 ----| 900	585	20.15	34	8.50
900 ----| 1000	317	10.92	23	5.75
> 1000	10	0.34	4	1.00
Without classification	3	0.10	-	-
Total	**2903**	**100**	**400**	**100**

In comparisons among the accessions sampled and the traditional collection, considering the similarity coefficient (CR%), phenotypic diversity coefficient (PD), and the homogeneity measured by the Shannon index (H´), the accessions sampled exhibited maximum similarity (100%) with the traditional collection. Values higher than 100% were found in various studies of composition of core collections using information of phenotypes in peanut and sweet potato crops (UPADHYAYA *et al*., 2002; HOLBROOK; DONG, 2005; VASCONCELOS *et al*., 2010). According to Hu *et al*. (2000), for the sample to be considered representative of the collection, the minimum limit for the similarity coefficient is 80.0%. The similarity coefficient measures the degree of agreement of the categories of variables and is directly related to the purpose of preservation of the alleles responsible for expression of the categories (phenotype) that were manifested in the collection. This reinforces the strategy of use of the multivariate models as a satisfactory strategy in establishing a core collection.

In relation to the phenotypic diversity coefficient, the accessions sampled had a response superior to the collection of traditional accessions, which has the value 1.00. The value was 1.14 for phenotypic data, 1.09 for ecogeographic data, and 1.11 for the combined phenotype-ecogeographic data. Values of PD > 1.0 indicate that the distribution of the accessions in the categories of the variables is more balanced in the set of accessions sampled than in the collection of traditional accessions (SOUZA; PEÑALOZA, 2005). For the Shannon diversity index (H’), the values obtained for the sampled accessions and the traditional collection were 1.74 and 1.47 for phenotypic data, 2.11 and 1.83 for ecogeographic data, and 1.95 and 1.68 for the combined phenotypic-geographic data, respectively. The greater the value of H', the greater the phenotypic and/or ecogeographic diversity of the population under study will be, expressing the richness and the uniformity in a core collection (ABADIE *et al*., 2005). The study of diversity through the phenotypic diversity coefficient and the Shannon diversity index complements the evaluation of the similarity coefficient. In this case, the study of diversity evaluates not only the representation of the categories of variables in the traditional collection, but also how much the distribution of the accessions in these categories is balanced, so as to minimize the probability of genetic erosion.

Although there are various strategies in the literature for forming core collections that are able to represent the base collection, the strategy by multivariate models represented the traditional collection very well in the accessions sampled. In all aspects, it exhibited similarity equal to the categories in relation to the phenotypic and ecogeographic descriptors of the traditional collection. In contrast, the phenotypic diversity and Shannon coefficients were greater than the coefficients in the traditional collection, showing greater distribution of phenotypic diversity in the group of accessions sampled than in the distribution in the traditional collection. However, this does not necessarily imply that all the alleles were included, since in practice, it would be impossible to include all the accessions in a single sample (VAN HINTUM *et al*., 2000). Consequently, the set of accessions formed by the phenotypic-geographic multivariate strategy composed of 400 traditional accessions (Figure 2b) well represented the base collection originating from the traditional collections, with a minimum of genetic erosion. For that reason, the proposal was used to allocate accessions for formation of the core collection.

The core collection that was formed represents 9.5% of the traditional accessions coming from the collections carried out in Brazil (Figure 1a). In relation to the base collection, the core collection proposed represented a frequency of 3%, greater than the frequency proposed by various authors (UPADHYAYA; ORTIZ, 2001; ZHANG *et al*., 2010).

Acknowledgments

We would especially like to thank Dr. Jaison Pereira de Oliveira (*in memoriam*); although he is no longer with us, he is the author of this article and continues to inspire by his example and dedication to science throughout his career.

We also express our gratitude to Dr. Michael Thung (*in memoriam*), a life-long enthusiast of beans research, who collaborated in building the core collection and reviewing this article.

References

ABADIE, T.; MAGALHÃES, J.R.; PARENTONI, S.; CORDEIRO, C.M.T.; ANDRADE, R. The Core Collection of Maize Germplasm of Brazil. **Plant**

Genetic Resources Newsletter, n.117, p.55-56, 1999.

ABADIE, T.; CORDEIRO, C.M.T.; FONSECA, J.R.; ALVES, R.B.N.; BURLE, M.L.; BRONDANI, C.; RANGEL, P.H.N.; CASTRO, E.M.; SILVA, H.T.; FREIRE, M.S.; ZIMMERMANN, F.J.P.; MAGALHÃES, J.R. Construção de uma coleção nuclear de arroz para o Brasil. **Pesquisa Agropecuária Brasileira**, v.40, n.2, p.129-136, 2005. https://doi.org/10.1590/S0100-204X2005000200005

BORSUK, L.J.; SAIFERT, L.; VILLAMIL, J.M.O.; MORA, F.D.S.; NODARI, R.O. Phenotypic variability in feijoa fruits [*Acca sellowiana* (O. Berg.) Burret] on indigenous lands, quilombolas communities and protected areas in the south of Brazil. **Revista Brasileira de Fruticultura**, v.39, n.1, p.1-10, 2017. http://dx.doi.org/10.1590/0100-29452017699

BURLE, M.L.; FONSECA, J.R.; KAMI, J.A.; GEPTS, P. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. **Theoretical and Applied Genetics,** v.121, n.1, p.801–813, 2010. https://doi.org/10.1007/s00122-010-1350-5

BUSSAB, W.O.; MIAZAKI, E.S.; ANDRADE, D.F. Introdução à análise de agrupamentos. *In*: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9. **Resumos […].** São Paulo: Associação Brasileira de Estatística, 1990. p.105.

CHANDRA, S.; HUAMAN, Z.; HARI KRISHNA, S.; ORTIZ, R. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data—a simulation study. **Theoretical and Applied Genetics,** v.104, n.8, p.1325-1334, 2002. https://doi.org/10.1007/s00122-001-0854-4

FRANCO, J.; CROSSA, J.; TABA, S.; SHANDS, H. A sampling strategy for conserving genetic diversity when forming core subsets. **Crop Science**, v.45, n.3, p.1035-1044, 2005. https://doi.org/10.2135/cropsci2004.0292

FRANCO, J.; CROSSA, J.; WARBURTON, M.L.; TABA, S. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. **Crop Science,** v.46, n.2, p.854-864, 2006. https://doi.org/10.2135/cropsci2005.07-0201

FRANKEL, O.H.; BROWN, A.H.D. Plant genetic resources today: A critical appraisal. *In*: HOLDEN, J.H.W.; WILLIAMS, J.T. (eds.). **Crop genetic resources:** conservation and evaluation. Massachusetts: Allen and Unwin, 1984.

GOUESNARD, B.; BATAILLON, T.M.; DECOUX, G.; ROZALE, C.; SCHOEN, D.J.; DAVID, J.L. MSTRAT: an algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. **Journal of Heredity,** v.92, n.1, p.93-94, 2001. https://doi.org/10.1093/jhered/92.1.93

GRIGOLO, S.; CAROLINA, A.; DENARDI, S.; VACARI, J. Implicações da análise univariada e multivariada na dissimilaridade de acessos de feijão comum. **Rev. Ciênc. Agrovet**, v.17, n.3, p.351-360, 2018. https://doi.org/10.5965/223811711732018351

HOLBROOK, C.C.; DONG, W. Development and evaluation of a mini core collection for the U.S. peanut germoplasm collection. **Crop Science**, v.45, n.4, p.1540-1544, 2005. https://doi.org/10.2135/cropsci2004.0368

HU, J.; ZHU, J.; XU, H.M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. **Theoretical and Applied Genetics**, v.101, n.1, p.264-268, 2000. https://doi.org/10.1007/s001220051478

JIANG, Q.; HOU, J.; HAO, C.; WANG, L.; GE, H.; DONG, Y.; ZHANG, X. The wheat (*T. aestivum*) sucrose synthase 2 gene (TaSus2) active in endosperm development is associated with yield traits. **Funct. Integr. Genomics**, v.11, n.1, p.49-61, 2010. https://doi.org/10.1007/s10142-010-0188-x

KIM, K.W.; CHUNG, H.K.; CHO, G.T.; MA, K.H.; CHANDRABALAN, D.; GWAG, J.G.; KIM, T.S.; CHO, E.G.; PARK, Y.J. PowerCore: a program applying the advanced m strategy with a heuristic search for establishing core sets. **Bioinformatics**, v.23, n.16, p.2155-2162, 2007. http://dx.doi.org/10.1093/bioinformatics/btm313

KNUPP, A.M.; FERREIRA, E.P.B.; ARAÚJO, A.P. Variability of nodulation traits in Andean and Mesoamerican common bean gene pools. **Pesq. Agrop. Bras.**, v.52, n.4, p.252-260, 2017. https://doi.org/10.1590/s0100-204x2017000400005

LI, C.T.; SHI, C.H.; WU, J.G.; XU, H.M.; ZHANG, H.Z.; REN, Y.L. Methods of developing core collections based on the predicted genotypic value of rice (*Oryza sativa* L.). **Theoretical and Applied Genetics**, v.108, n.6, p.1272-1176, 2004. https://doi.org/10.1007/s00122-003-1536-1

LIU, W.; SHAHID, M. Q.; BAI, L.; LU, Z.; CHEN, Y.; JIANG, L.; DIAO, M.; LIU, X.; LU, Y. Evaluation of Genetic Diversity and Development of a Core Collection of Wild Rice (*Oryza rufipogon* Griff.) Populations in China. **PLoS One**, v.10, n.12, p.1-23, 2015. https://doi.org/10.1371/journal.pone.0145990

MARCHINI, R.M.S. **Caracterização de acessos do "complexo Saccharum" e desenvolvimento de uma coleção nuclear com base em marcadores moleculares e atributos de qualidade**. 2019. Tese (Doutorado) - Faculdade de Ciências Agrárias e Veterinárias – Unesp, Jaboticabal, 2019.

NASS, L.L. **Recursos genéticos vegetais**. Brasília: Embrapa, 2007.

NAYAK, S.N.; SONG, J.; VILLA, A.; PATHAK, B.; AYALA-SILVA, T.; YANG, X.; TODD, J.; GLYNN, NC.; KUHN, D.N.; GLAZ, B.; GILBERT, R.A.; COMSTOCK, J.C.; WANG, J. Promoting Utilization of Saccharum spp. Genetic Resources through Genetic Diversity Analysis and Core Collection Construction. **PLoS One**, v.9, n.10, p.1-12, 2014. https://doi.org/10.1371/journal.pone.0110856

ODONG, T.L.; JANSEN, J.; EUWIJK, F.A.V.; HINTUM, T.J.L.V. Quality of core collections for effective utilization of genetic resources review, discussion and interpretation. **Theor. Appl. Genet,** v.126, n.2, p.289-305, p.2013. https://doi.org/10.1007/s00122-012-1971-y

POLICRITI, A.; SGARRO, A. **CoreFinder v.1.0**. 2011.

RANGEL, P.H.N.; OLIVEIRA, P.J.; COSTA, J.G.C.; FERREIRA, M.E.; ABREU, A.G. **Banco ativo de germoplasma de arroz e feijão:** passado, presente e futuro. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2013. 68 p.

SHANNON, C.E.; WEAVER, W. **The mathematical theory of communication**. Urbana: University of Illinois Press, 1949.

SHARMA, M.; RATHORE, A.; MANGALA, U.N.; GHOSH, R.; SHARMA, S.; UPADHYAY, H.; PANDE, S. New sources of resistance to Fusarium wilt and sterility mosaic disease in a mini-core collection of pigeonpea germplasm. **Eur J. Plant. Pathol,** v.133, n.3, p.707-714, 2012. https://doi.org/10.1007/s10658-012-9949-9

SILVA, A.S.L. **Diversidade genética entre genótipos de soja e estabelecimento de coleções nucleares e de melhoramento para teores de óleo e de proteína**. 2016. Dissertação (Mestrado) - Universidade Federal de Viçosa, Viçosa, 2016.

SILVA, C.C.; RODRIGUES, D.P.; CLEMENT, C.R.; ASTOLFI FILHO, S. Molecular-genetic analysis for validation of peach palm (*Bactris gasipaes* Kunt) landraces using RAPD markers. **Científica,** v.47, n.3, p.313-320, 2019. https://doi.org/10.15361/1984-5529.2019v47n3p313-320

SILVA, R.N.O.; BURLE, M.L.; PADUA, J.G.; LOPES, A,C,A.; GOMES, R. L. F.; MARTINEZ-CASTILLO, J. Phenotypic diversity in lima bean landraces cultivated in Brazil, using the Ward-MLM strategy. **Chil. J. Agr. Res.**, v.77, n.1, p.35-40, 2017. https://doi.org/10.4067/S0718-58392017000100004

SOUZA, J.; PEÑALOZA, R.A.S. **Estatística exploratória de séries nominais e ordinais: teoria de aplicação a medidas multidimensionais nebulosas de pobreza para dados ordinais**. Campinas: Mimeo, 2005.

UPADHYAYA, H.D.; BRAMEL, P.; ORTIZ, R.; SINGH, S. Developing a mini core of peanut for utilization of genetic resources. **Crop Science**, v.42, n.6, p.2150-2156, 2002. https://doi.org/10.2135/cropsci2002.2150

UPADHYAYA, H.D.; ORTIZ, R. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. **Theoretical and Applied Genetics**, v.102, n.8, p.1292-1298, 2001. https://doi.org/10.1007/s00122-001-0556-y

UPADHYAYA, H.D.; SARMA, N.D.R.K.; RAVISHANKAR, C.R.; ALBRECHT, T.; NARASIMHUDU, Y.; SINGH, S.K.; VARSHNEY, S.K.; REDDY, V.G.; SINGH, S.; DWIVEDI, S.L.; WANYERA,

N.; ODUORI, C.O.A.; MGONJA, M.A.; KISANDU, D.B.; PARZIES, H.K.; GOWDA, C.L.L. Developing a Mini-Core Collection in Finger Millet Using Multilocation Data. **Crop Science**, v.50, n.5, p.1924-1931, 2010. https://doi.org/10.2135/cropsci2009.11.0689

VALDISSER, P.A.M.R.; PEREIRA, W.J.; ALMEIDA FILHO, J.E.; MÜLLER, B.S.F.; COELHO, G.R.C.; MENEZES, I.P.P.; VIANNA, J.P.G.; ZUCCHI, M.I.; LANNA, A.C.; COELHO, A.S.G.; OLIVEIRA, J.P.; MORAES, A.C.; BRONDANI, C.; VIANELLO,R.P. In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. **Bmc Genomics,** v.18, n.1, p.1-19, 2017. https://doi.org/10.1186/s12864-017-3805-4

VASCONCELOS, E.D.; CRUZ, C.D.; BHERING, L.P.; FERREIRA, A. Estratégias de amostragem e estabelecimento de coleções nucleares. **Pesquisa agropecuária brasileira,** v.42, n.4, p.507-514, 2007. https://doi.org/10.1590/S0100-204X2007000400008

VASCONCELOS, E.S.; CRUZ, C.D.; SILVA, D.J.H.; BHERING, L.L. Tamanho de coleção original, métodos de agrupamento e amostragem para obtenção de coleção nuclear de germoplasma. **Pesquisa agropecuária brasileira**, v.45, n.12, p.1448-1455, 2010. https://doi.org/10.1590/S0100-204X2010001200016

WANG, J.; GUAN, Y.; WANG, Y.; ZHU, L.; WANG, Q.; HU, Q.; HU, J. A Strategy for Finding the Optimal Scale of Plant Core Collection Based on Monte Carlo Simulation. **The Scientific World Journal**, v.2014, n.1, p.1-9, 2014. http://dx.doi.org/10.1155/2014/503473

ZEWDIE, Y.; TONG, N.K.; BOSLAND, P. Establishing a core collection of Capsicum using a cluster analysis with enlightened selection of accessions. **Genetic Resources and Crop Evolution,** v.51, n.2, p.147-151, 2004. https://doi.org/10.1023/B:GRES.0000020858.96226.38

ZHANG, H.; ZHANG, D.; WANG, M.; SUN, J.; QI, Y.; LI, J.; WEI, X.; HAN, L.; QIU, Z.; TANG, S.; LI, Z. A core collection and mini core collection of *Oryza sativa* L. in China. **Theoretical and Applied Genetics**. v.122, n.1, p.49-61, 2010. https://doi.org/10.1007/s00122-010-1421-7