

Population structure and genetic diversity in common bean accessions from Northeast Brazil

Maria da Conceição Martiniano-Souza ^{1,*}, Maria Celeste Gonçalves-Vidigal ², Giseli Valentini ², Julio Cesar Ferreira Elias ², Larissa Fernanda Segal Xavier ², Andrea Ariani ³, Paul Gepts ³ and Antonio Félix da Costa ¹

¹ Department of Agricultural Research, Agronomic Institute of Pernambuco, Recife, Pernambuco, Brazil.

² Department of Agronomy, Maringa State University, Maringa, Parana, Brazil.

³ Department of Plant Sciences, Section of Crop and Ecosystem Sciences, University of California, Davis, CA, United States of America.

World Journal of Advanced Research and Reviews, 2021, 12(01), 287–297

Publication history: Received on 14 September 2021; revised on 15 October 2021; accepted on 17 October 2021

Article DOI: <https://doi.org/10.30574/wjarr.2021.12.1.0532>

Abstract

Common bean is of great relevance for several countries, especially those located in tropical regions. Common bean in Brazil has a wide genetic diversity, which is used by breeders to improve this crop. In this study, we used GBS methodology to genotype common bean accessions from Northeast Brazil, and to study its population structure and genetic diversity. After a filtering process, we identified 30,529 high-quality SNPs distributed in 11 linkage groups. The number of SNPs per chromosome ranged from 1,731 to 3,853. The population structure analysis separated the accessions into two subpopulations, according to $K=2$, one subpopulation with 26 Andean and the other with 60 Middle American accessions. Considering $K=3$, Middle American accessions were divided into two subpopulations with presence of allele mixture between these two groups. Based on the PCA, we were able to verify a narrow genetic base of accessions belonging to the Andean gene pool, as well as a vast genetic diversity among the accessions from the Middle American gene pool. The knowledge on the genetic diversity among the accessions is of extreme importance to subsidize the common bean breeding programs in Northeast Brazil, and to explore the variability existing in cultivars adapted to the specific bioclimatic conditions.

Keywords: Genetic variability; Genotyping-by-sequencing; Molecular analysis; *Phaseolus vulgaris*; Principal component analysis; Single nucleotide polymorphism

1. Introduction

The common bean (*Phaseolus vulgaris* L.) is the main legume for direct human consumption in the world, especially in developing countries in Africa and Latin America, due its great amount of proteins, complex carbohydrates, fiber and essential vitamins [1]. Brazil is the largest global producer of dry beans, followed by Mexico and the United States. Common bean production in Brazil has been relatively constant at around 3 million metric tons over the years, as observed in the production for the 2019/2020 agricultural year, distributed in three different crop seasons, that was 3.2 million metric tons cultivated in an area of 2.9 million hectares [2].

The Northeast region accounts 26.6% of the common bean production in Brazil, yielding around 0.568 t ha⁻¹, which is far lower than the national yield of 1.1 t ha⁻¹. In contrast, the Northeast cultivated area extends 1.5 million hectares, corresponding to 51% of the total cultivated area in Brazil [2]. Low yields are due in part to the level of technology employed, but it is, undoubtedly, more correlated with the impact of droughts, one of the main abiotic factors limiting

* Corresponding author: Maria da Conceição Martiniano-Souza; E-mail: mariamartiniano@hotmail.com
Department of Agricultural Research, Agronomic Institute of Pernambuco, Recife, Pernambuco, Brazil.

common bean production worldwide [3, 4]. The Brazilian semi-arid region comprises part of the northeast and the north of Minas Gerais, considering only the Brazilian northeast, approximately 64.82% is recognized as a semi-arid land [5]. The annual rainfall in this region can fluctuate from 2000 mm to values less than 500 mm in some areas, however there is an unfavourable distribution of rainfall through the crop seasons [6]. Future climate projections suggest that the occurrence of drought would be more frequent, with increase in temperature and reduction in rainfall [6, 7]. In this context, the genetic diversity available in the bean germplasm will play an important role to reduce drought effect on the common bean production. Breeding programs can take advantage of the genetic variability existent in local varieties to develop more resilient cultivars adapted to the Northeast specific climatic condition.

Common bean is produced in all Brazilian states, and its cultivation is characterized either by subsistence or high input agriculture. The two most cultivated and consumed common bean market classes are Carioca and Black beans [8]. In addition, other market classes of beans, such as Favita, Jalo, Rosinha, Mulatinho, and others, have their cultivation restricted to smaller areas [9]. The common bean germplasm, especially landraces or traditional cultivars that have been cultivated for several years at the same location, present broad phenotypical and genotypical variability, and are adapted to specific ecological conditions; they can be also resistant to pests and diseases, characteristics of great importance for common bean breeding programs aiming to develop cultivars with enhanced attributes [10].

The domestication of the common bean occurred independently in two different regions of the Americas [11, 12, 13, 14]. Cultivars with small seeds are originally from Central America, from northern Mexico to Colombia, and belong to the gene pool known as Middle American. Beans with large seeds belong to the Andean gene pool and are originally from South American countries as Peru, Chile, Bolivia, and Argentina [15]. These two major ecogeographic gene pools were recognized through different characteristics, such as morphological traits, type of phaseolin (main protein reserve in seeds of cultivated and wild beans), variation of mitochondrial DNA [16], isoenzymes and molecular markers [17, 15, 18].

For the common bean breeding program, it is fundamental to establish, preserve and use a germplasm collection of traditional or improved cultivars of common beans from both Middle American and Andean gene pools, as representatives of the genetic diversity in this crop. Also, an efficient conservation and use of the existing germplasm is a crucial prerequisite to decipher the population structure of these accessions, and to build a report on the genetic diversity distribution and its relation to the geographical and ecological information about this specific region [12, 19].

Molecular biology has largely contributed to develop effective and oriented breeding strategies, allowing predictions of relationships among cultivars, identification of promising accessions for hybridization, and identification of duplicates [20]. The development of effective and tightly linked molecular markers for several traits of interest also contributes for the selection of genotypes through marker-assisted selection (MAS) [21, 22]. Through genotyping-by-sequencing (GBS), thousands of single nucleotide polymorphisms (SNPs) are obtained from a group of individuals [23]. The identification of SNPs allows to understand the genetic diversity in the germplasm, and it provides a source of variations for further association studies of characteristics of interest, and for the selection of bean accessions adapted to specific environmental conditions [24]. Therefore, the objective of this study was to evaluate the population structure and genetic diversity of common bean accessions from the Northeast region of Brazil, using single nucleotide polymorphism (SNP) markers obtained by genotype-by-sequencing (GBS).

2. Material and methods

This study was conducted at the Nucleo de Pesquisa Aplicada a Agricultura (Nupagri), at the State University of Maringá (UEM), Maringá, Paraná, Brazil, and the Laboratory of Plant and Environmental Sciences, at the University of California Davis, California, USA. A total of 86 accessions of common beans evaluated in this study are from the Northeast region of Brazil (Table 1).

Among the accessions, 31 belong to the Common Bean Germplasm Bank of the Instituto Agronômico de Pernambuco (BAG-Bean-IPA) and 55 accessions are traditional cultivars collected in bean production areas located in 22 municipalities in the state of Pernambuco, in Northeast Brazil (Table 2). The common bean accessions are from the Agreste (n=69), Sertão (n=12), Zona da Mata (n=1), and Metropolitana (n=4) mesoregions of the state of Pernambuco. The 86 accessions belong to the market classes Carioca, Black, Mulatinho, or Others. The “Others” market class comprises several accessions that belong to different subgroups: White, Jalo, Rosinha, Purple, Jalinho, Red, Rajado, Pintado, Enxofre and Pardo.

Table 1 Accession codes, common names of common bean accessions, market classes of common bean accessions, municipalities where common bean accessions were collected in the Northeast region of Brazil

Code	Common name	Market class	Gene Pool	Municipality
BL-1	Brígida	Carioca	MA	Recife, PE ⁴
BL-2	Cocão	Others	A	Recife, PE
BL-3	Bagajó	Others	A	Poço Verde, SE ⁵
BL-4	Favita	Others	A	Garanhuns, PE
BL-5	Canarinho	Others	A	Lajedo, PE
BL-6	Rosinha claro	Others	MA	Calçado, PE
BL-7	Chita Fina Verdadeira ¹	Others	A	São João, PE
BL-8	Jaula	Others	A	Garanhuns, PE
BL-9	Pintado	Others	A	Ibimirim, PE
BL-10	Balinha	Mulatinho	MA	Lajedo, PE
BL-11	Praia	Others	A	Poço Verde, SE
BL-12	Camarão	Others	A	Calçado, PE
BL-13	BSF-1	Others	A	BSF ¹ , PE
BL-14	BSF-2 (Pingo de Ouro)	Carioca	MA	BSF ¹ , PE
BL-15	BSF-3 (Fogo na Serra)	Others	A	BSF ¹ , PE
BL-199	Enxofre	Others	A	Lajedo, PE
BL-18	Africano 4	Others	A	Recife, PE
BL-19	IPA 1	Mulatinho	MA	Recife, PE
BL-222	Gordo	Others	A	Lajedo, PE
BL-223	Mulatinho	Mulatinho	MA	Águas Belas, PE
BL-224	BRS-Marfim	Mulatinho	MA	São João, PE
BL-25	Jucati 4	Mulatinho	MA	Jucati, PE
BL-27	Mulatão	Others	A	Bezerros, PE
BL-28	Bairosa 2	Mulatinho	MA	Lajedo, PE
BL-29	Milagre de Santo Antônio	Mulatinho	MA	Águas Belas, PE
BL-30	Flor Azul	Mulatinho	MA	Águas Belas, PE
BL-31	Bico de Ouro	Mulatinho	MA	Águas Belas, PE
BL-33	HF-465-63-1	Mulatinho	MA	Caruaru, PE
BL-34	Laje	Mulatinho	MA	SMT ² , PB ⁶
BL-225	Sempre Assim	Others	MA	Águas Belas, PE
BL-226	CLPE3	Black	MA	São João, PE
BL-227	CLPE4	Black	MA	São João, PE
BL-41	CLPE7	Black	MA	São João, PE
BL-228	CLPE8	Black	MA	São João, PE
BL-C	CLPE9	Carioca	MA	São João, PE
BL-229	CLPE10	Black	MA	Caetés, PE
BL-230	CLPE11	Black	MA	Jucati, PE
BL-231	CLPE12	Mulatinho	MA	São João, PE
BL-232	CLPE14	Carioca	MA	São João, PE
BL-233	CLPE15	Mulatinho	MA	Caetés, PE
BL-M	CLPE16	Mulatinho	MA	Arcoverde, PE
BL-50	CLPE17	Black	MA	Lajedo, PE

BL-234	CLPE21	Black	MA	Calçado, PE
BL-57	CLPE25	Black	MA	Lajedo, PE
BL-216	CLPE32	Carioca	MA	SMC ³ , PE
BL-0	CLPE35	Outros	A	Arcoverde, PE
BL-65	CLPE37	Black	MA	Jucati, PE
BL-66	CLPE40	Carioca	MA	Caruaru, PE
BL-67	CLPE41	Carioca	MA	SMC ³ , PE
BL-68	CLPE43	Mulatinho	MA	Jupi, PE
BL-69	CLPE44	Mulatinho	MA	Arcoverde, PE
BL-70	CLPE45	Carioca	MA	Vertente, PE
BL-71	CLPE47	Mulatinho	MA	Arcoverde, PE
BL-72	CLPE50	Others	A	Arcoverde, PE
BL-73	CLPE52	Others	A	Arcoverde, PE
BL-74	CLPE53	Others	A	São João, PE
BL-75	CLPE54	Others	A	São João, PE
BL-76	CLPE55	Carioca	M	São João, PE
BL-77	CLPE56	Others	A	São João, PE
BL-78	CLPE58	Others	A	São João, PE
BL-79	CLPE60	Others	A	São João, PE
BL-80	CLPE61	Black	MA	São João, PE
BL-81	CLPE63	Black	MA	São João, PE
BL-82	CLPE65	Carioca	MA	São João, PE
BL-83	CLPE66	Black	MA	Lajedo, PE
BL-84	CLPE67	Black	MA	Jucati, PE
BL-85	CLPE68	Carioca	MA	Jucati, PE
BL-86	CLPE69	Black	MA	Jucati, PE
BL-87	CLPE74	Carioca	MA	Arcoverde, PE
BL-88	CLPE75	Mulatinho	MA	São João, PE
BL-89	CLPE79	Mulatinho	MA	Jupi, PE
BL-90	CLPE80	Black	MA	São João, PE
BL-91	CLPE81	Black	MA	São João, PE
BL-92	CLPE87	Mulatinho	MA	Arcoverde, PE
BL-93	CLPE88	Others	A	Casinhas, PE
BL-94	CLPE85	Others	A	Lajedo, PE
BL-95	CLPE90	Black	MA	Calçados, PE
BL-96	CLPE91	Mulatinho	MA	Caruaru, PE
BL-98	CLPE93	Others	A	Caruaru, PE
BL-99	CLPE92	Black	MA	Caruaru, PE
BL-100	CLPE89	Mulatinho	MA	Calçados, PE
BL-101	CLPE84	Mulatinho	MA	Riacho das Almas, PE
BL-102	CLPE94	Black	MA	São Caitano, PE
BL-103	CLPE96	Black	MA	São Caitano, PE
BL-104	CLPE86	Others	MA	Surubim, PE
BL-105	CLPE83	Mulatinho	MA	SMC ³ , PE

¹BSF - Belém do São Francisco; ²SMT - São Miguel de Taipu; ³SMC - Santa Maria do Cambucá; ⁴PE - Pernambuco; ⁵SE - Sergipe; ⁶PB - Paraíba

Table 2 Geographical description of 22 municipalities in Northeast, Brazil where 86 accessions of common bean *Phaseolus vulgaris* were collected for this study

Municipality	State	Latitude (S)	Longitude (W)	Altitude (masl)	Num. Accessions	Mesoregion
Casinhas	PE	07°44'28"	35°43'16"	390	1	Agreste
Surubim	PE	07°49'55"	35°45'21"	394	1	Agreste
SMC ¹	PE	07°50'24"	35°54'07"	494	3	Agreste
Vertente	PE	07°54'10"	35°59'16"	401	1	Agreste
RA ²	PE	08°08'02"	35°51'21"	407	1	Agreste
Caruaru	PE	08°16'58"	35°58'33"	554	5	Agreste
São Caitano	PE	08°19'33"	36°08'34"	552	2	Agreste
Lajedo	PE	08°39'50"	36°19'12"	661	9	Agreste
Jucati	PE	08°42'21"	36°29'20"	820	6	Agreste
Jupi	PE	08°42'43"	36°24'54"	782	2	Agreste
Calçado	PE	08°44'31"	36°20'02"	643	5	Agreste
Caetés	PE	08°46'22"	36°37'22"	849	2	Agreste
São João	PE	08°52'33"	36°22'01"	716	21	Agreste
Bezerros	PE	08°53'24"	36°29'34"	470	1	Agreste
Garanhuns	PE	08°53'25"	36°29'34"	842	2	Agreste
Águas Belas	PE	09°06'39"	37°07'22"	376	5	Agreste
Poço Verde	SE	10°42'28"	38°10'58"	268	2	Agreste
Recife	PE	08°03'14"	34°52'51"	4	4	Metropolit. ³
Arcoverde	PE	08°25'15"	37°03'41"	663	8	Sertão
Ibimirim	PE	08°32'26"	37°41'25"	401	1	Sertão
BSF ⁴	PE	08°45'28"	38°57'50"	305	3	Sertão
SMT ⁵	PB	07°15'00"	35°12'36"	45	1	ZM ⁶

¹ SMC - Santa Maria do Cambucá; ² RA - Riacho das Almas; ³ Metropolit. - Metropolitana; ⁴ BSF - Belém do São Francisco; ⁵ SMT - São Miguel de Taipu; ⁶ ZN - Zona da Mata

The 86 common bean accessions were genotyped through genotyping-by-sequencing (GBS). The seeds of each accession were seeded in plastic trays (48 × 30 × 11 cm) containing peat-based substrate (Humusfértil® Toledo, Parana, Brazil) and they were kept in the greenhouse until the plants presented their first trifoliolate leaves, which were then collected for DNA extraction. The DNA was extracted from freeze-dried leaves using the protocol proposed by Pallotta *et al.* [25], with modifications. The DNA quality was measured using NanoDrop Lite (Thermo Fisher Scientific/Invitrogen, Grand Island, NY) and by 1 percent agarose gel electrophoresis. DNA samples were considered to be high quality when the absorbance ratio (A260/A280) was higher than 1.8 and when the samples did not show visible degradation in agarose gel. Then, the DNA samples were used for subsequent library preparation. Genomic DNA and library adapters (Barcodes) were quantified with Qubit dsDNA HS assay kit (Thermo Fisher Scientific/Invitrogen, Grand Island, NY).

The GBS methodology was used to perform the genotyping of accessions, following the protocol described by Ariani *et al.* [24]. The restriction enzyme used for DNA digestion was CviAII (New England Biolabs, Ipswich, MA), which recognizes the C'ATG site. The genotyping of each accession went through the processes of *in silico* digestion, preparation of libraries and sequencing, as described by Ariani *et al.* [24].

After obtaining the genotyped products, the subsequent steps involved sequencing pre-processing, alignment of sequences with the reference genome of *P. vulgaris*, cultivar G19833, and SNP calling, following the methodology described by Ariani *et al.* [24]. A SNP was selected for further analysis when the following requirements were: minor

allele frequency (MAF) > 0.05; minimum quality > 10; and a mean read depth, across all lines, ranging from 5 to 1000 (-maf 0.05-minQ 10 -min-meanDP 5 -max-meanDP 1000).

SNPs located in repetitive regions were removed with VCFtools, using the reference genome annotation of *P. vulgaris* version 1.0 available at Phytozome. Only SNPs with DNA coding sequences were used. Accessions that exhibited <10 percent quality, missing data and heterozygous profile (allele frequency higher than 0.05) were not included in the phylogenetic analysis. Finally, an excel file was constructed containing the filtered SNPs extracted from each position for each 86-bean accession. The final data set was used for the phylogenetic analysis and to compare the population structure and levels of genetic diversity between accessions.

2.1. Statistical Analysis

The methodology followed for the phylogenetic analysis was proposed by Ariani *et al.* [24]. The phylogenetic tree was built based on Neighbor-Joining Tree (NJ) clustering methodology, considering the Kimura two-parameter model with 1000 bootstrap replicates. The phylogenetic analysis was conducted using MEGA7 software [26]. The colour of each accession in the Neighbor-Joining Tree was based on the bean market classes Mulatinho, Carioca, Black and Others.

In order to evaluate the genetic structure of the 86 common bean accessions from Northeast Brazil, we performed the classic Bayesian population structure analysis using LEA package, an R package for landscape and ecological association studies. Parameters for the population structure analysis included the number of K from 2 to 10 groups and 10 interactions, as described by Frichot and François [27]. The Delta K parameter used to calculate and define the best number of subpopulations was obtained from LEA package.

The principal component analysis (PCA) was estimated using the default correlation method of TASSEL 5.0 software [28]. The values of the first two main components (PC1 and PC2) were used to build a two-dimensional graph containing the 86 bean accessions.

3. Results and discussion

The sequencing analysis via GBS generated a total of 30,529 SNPs distributed across 11 linkage groups and 5 scaffolds of *P. vulgaris* genome (Table 3). The highly diverse set of bean accessions from Northeast Brazil measured through the 30,529 SNPs suggests the presence of population structure.

Table 3 Linkage group, SNPs and number of SNPs (their positions in parentheses) identified for the 86 common bean accessions in Pernambuco, obtained with R package LEA (Landscape and Ecological Associations)

Linkage Group	SNPs	Number of SNPs (bp position)
Pv01	3,180	0 (183,346) – 3,179 (52,020,722)
Pv02	3,853	3,180 (17,813) – 7,032 (48,686,602)
Pv03	3,760	7,033 (272,759) – 10,792 (52, 272,872)
Pv04	1,922	10,793 (766,583) – 12,714 (45,037,645)
Pv05	2,339	12,715 (819,316) – 15,053 (40,643,363)
Pv06	2,330	15,054 (17,234) – 17,383 (31,956,823)
Pv07	2,464	17,384 (406,862) – 19,847 (51,743,071)
Pv08	2,518	19,848 (17,125) – 22,365 (59,493,010)
Pv09	3,547	22,366 (257,610) – 25,912 (37,392,701)
Pv10	1,731	25,913 (713,570) – 27,643 (43,187,875)
Pv11	2,861	27,644 (42,424) – 30,504 (50,209,148)
SCAFFOLD_12	1	30,505 (1,764) – 30,505 (1,764)
SCAFFOLD_13	3	30,506 (18,026) – 30,508 (18,065)
SCAFFOLD_15	7	30,509 (101,126) – 30,515 (228,763)
SCAFFOLD_17	6	30,516 (3,040) – 30,521 (14,171)
SCAFFOLD_23	7	30,522 (95,368) – 30,528 (140,660)
Total of SNPs	30,529	

According to Delta $K=2$ the accessions are divided into two subpopulations, representing the respective Andean and Middle American gene pools. One subpopulation consisted of 26 accessions belonging to the Andean gene pool and the second subpopulation comprises the 60 accessions of the Middle American gene pool (Figure 1). However, we were able to observe a mix of alleles between the two gene pools at a small rate, which is a result of the incorporation of alleles of the Andean gene pool in Middle American germplasm, as well as alleles of Middle American origin incorporated in Andean germplasm.

Based on these results, we highlight the fact that both Andean and Middle American common beans are being produced in the Northeast region of Brazil. Similar results were reported by Burle *et al.* [29] and Blair *et al.* [30] in studies involving common bean accessions from different regions in Brazil. Evaluating the phaseolin content in bean seeds, Burle *et al.* [29] demonstrated that 79 percent of the studied accessions had “S” phaseolin, from Middle American gene pool, and 21 percent had “T” phaseolin, from Andean gene pool, demonstrating that both common bean gene pools are being grown in the Brazilian territory. Common beans from the Middle American gene pool are the most cultivated in Brazil [31, 32, 29] and correspond to approximately 84 percent of common beans cultivated throughout Latin America [33].

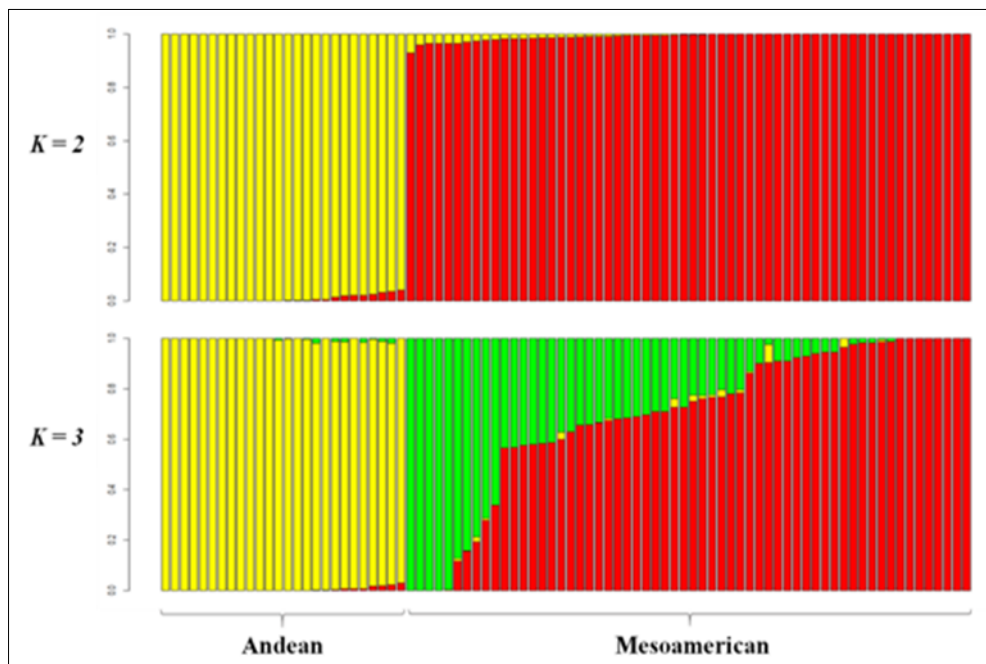


Figure 1 Population structure analysis for 86 common bean accessions from the Northeast region of Brazil, based on 30,529 SNPs obtained by GBS method. The number of subpopulations range from $K=2$ to $K=3$. Population Structure analysis was performed with R package LEA (Landscape and Ecological Associations)

At $K=3$, we observed the separation of Middle American accessions into two subpopulations, with considerable admixture among individuals (Figure 1). This fact reflects a large genetic variability within the Middle American germplasm. The Middle American group corresponds to 60 accessions that have small seeds, a characteristic of this gene pool, and are mainly represented by the market classes Carioca, Black and Mulatinho. These marker classes are broadly cultivated in the Brazilian Northeast region. Burle *et al.* [29] identified the admixture between accessions from Middle American subpopulations and a low frequency of allele introgression between gene pools. However, Blair *et al.* [30] reported considerable admixture between Andean and Middle American gene pools, analysing 362 common bean accessions from Brazil. Genetic flow is an important characteristic of the genetic structure, also occurring among common bean races, as it was evidenced by Blair *et al.* [34].

The relationship among accessions studied through Principal Component Analysis (PCA), genotyped with 30,529 SNPs, revealed the PC1 component separated the accessions between the Andean and Middle American gene pools, due to the great diversity observed between them (Figure 2). Principal Component, PC2 provided the identification of diversity within the Middle American gene pool, which has greater genetic variability. The first main component explains 90.8 percent of the variation between accessions and it is responsible for the separation of accessions in Middle American and Andean gene pools. The second main component explains 1.2 percent of the total variation. Together, the two main components explain 92 percent of the total variation.

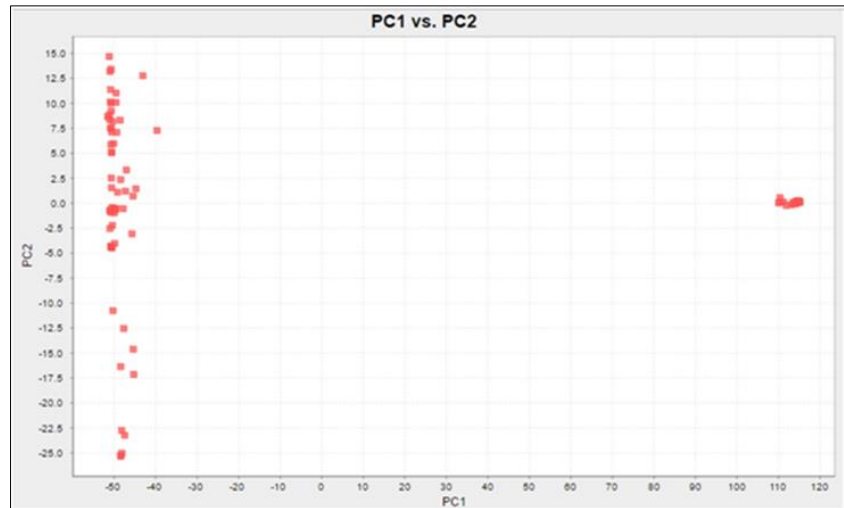


Figure 2 Representation of the two first Principal Components (PC1 vs. PC2) for 86 accessions of common beans from the Northeast region of Brazil, based on 30,529 SNPs. Principal component analysis was estimated with software Tassel 5.0 [28]

Middle American accessions have a high genetic variability in relation to Andean accessions, as shown by the high dispersion of points in the two-dimensional plane, corroborating the data reported by Singh *et al.* [15], Bitocchi *et al.* [12, 13], and Ariani *et al.* [24]. The lower genetic variability observed within the Andean gene pool could be related to the fact that a few samples of Middle American beans were the source of the Andean gene pool [35]. In addition, the domestication process of Andean common beans was characterized by a vast reduction in genetic variability [12], which can be perceived in the cultivated germplasm.

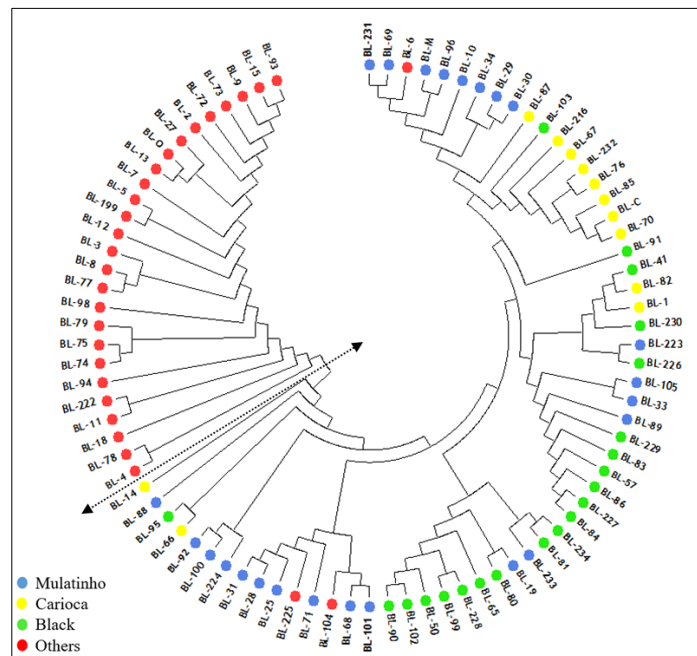


Figure 3 Neighbor-Joining Tree for 86 common bean accessions from the Northeast region of Brazil, based on 30,529 SNPs obtained by the GBS method. Neighbor-Joining Tree was constructed using MEGA7 software [26]. The color of each accession on the Neighbor-Joining Tree was based on the bean market classes Mulatinho, Carioca, Black and Others

The Neighbor-Joining Tree shows that accessions could be grouped efficiently according to the gene pool and market class (Figure 3). The 26 accessions from the Andean pool and previously classified as Other market class, were grouped in a main cluster. The Andean accessions have large seeds and are mainly from Enxofre and Favita market classes.

Enxofre and Favita beans belong to the Durango or Nova Granada races, which are from the Andean gene pool. The Neighbor-Joining Tree also shows that Middle American bean accessions were subdivided into subgroups (Figure 3).

Two main subgroups were formed, one containing Black bean accessions, and the other predominantly Carioca accessions. Mulatinho accessions were divided into two subclusters. Three accessions, BL-6, BL-104, and BL-225, belonging to the market classes Rosinha, Rajado and White, respectively, were classified in this study as Others. However, these accessions belong to the Middle American gene pool and they were grouped with the Middle American accessions in the Neighbor-Joining Tree (Figure 3).

Most of the common bean accessions cultivated in Northeast Brazil belong to the Middle American gene pool, due to the consumer preferences for small Carioca and Black beans. In addition, common bean consumers from Northeast Brazil also consume Mulatinho beans, which have historically been present in the daily diet of people living in that area. Mulatinho beans are Mesoamerican beans widely cultivated and consumed in the Northeast region of Brazil [30]. Typically, the Northeast region has a climate characterized by periods of high temperature and low precipitation, therefore, the Mulatinho cultivars demonstrate stability in the environment of the Northeast.

The knowledge of the genetic variability of common bean accessions cultivated by Northeastern farmers and accessions belonging to BAG-Bean-IPA is of great importance for common bean breeding programs for Northeast Brazil. This knowledge provides effectiveness and security in *P. vulgaris* germplasm conservation of Andean and Middle American gene pools and it allows the selection of potential combinations between accessions to generate genetic variability to develop new varieties. In addition, the data reveals that the accessions kept at BAG-Bean-IPA and by common bean producers represent a very diverse set of alleles and genes that could be useful for developing common bean cultivars adapted to the Northeast region of Brazil.

In the state of Pernambuco, for example, the Agreste mesoregion, located between Zona da Mata (wet region) and Sertão (dry region) concentrates the largest area of common bean production. The water supply and mild temperatures in the Agreste mesoregion provide edaphoclimatic conditions that favor the cultivation of beans. On the other hand, in the Zona da Mata and Metropolitana regions the excessive humidity reduces the common bean productive potential, while in the Northeastern Sertão, water scarcity is a limiting factor for bean production [4]. It is expected that the common beans cultivated in these different biomes show specific adaptive characteristics for the environment where they are cultivated, and therefore they could be used as source of alleles for breeding programs to develop cultivars adapted to specific environments.

This is the first report of a study involving common bean accessions from the Northeast region of Brazil and the use of SNP markers discovered by genotyping-by-sequencing (GBS) technique. The GBS data will be useful for future studies, for association analyses between the genotyped accessions and their relationship with resistance to biotic and abiotic factors. The proposal is to increase the knowledge on each accession and provide remarkable information for common bean breeding programs in Northeast Brazil.

4. Conclusion

The molecular analysis using the GBS technique showed that 86 common bean accessions from Pernambuco belong to Middle American and Andean gene pools. Middle American accessions have higher genetic diversity when compared to Andean accessions. This technique allowed the identification of 30,529 SNPs distributed in 11 linkage groups of common beans. The knowledge on the genetic diversity between these accessions is extremely important to subsidize the genetic improvement programs for common beans in the Northeast region of Brazil, allowing to explore the existing variability in cultivars already adapted to the climate conditions of the region.

Compliance with ethical standards

Acknowledgments

The authors gratefully acknowledge the Coordination for the Improvement of Higher Education Personnel (CAPES) and the National Council for Scientific and Technological Development (CNPq) for the financial support, and the team at the common bean program of Agronomic Institute of Pernambuco (IPA) for collecting the common bean samples used in this study. A.A. was supported by Agriculture and Food Research Initiative (AFRI) Competitive Grant N° 2013-67013-21224 from the USDA National Institute of Food and Agriculture to P.G.

Disclosure of conflict of interest

All the authors declare that there is no conflict of interest.

References

- [1] Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J. Beans (*Phaseolus* spp.): model food legumes. *Plant and Soil*. 2003; 252(1):55-128.
- [2] Conab-Companhia Nacional de Abastecimento. Feijão Total (1ª, 2ª e 3ª Safras). [cited 2021 April 21].
- [3] Beebe SE, Rao IM, Blair MW, Acosta-Gallegos AA. Phenotyping common beans for adaptation to drought. *Frontiers in physiology*. 2013; 4: 35.
- [4] Costa AF, Lopes LHO. Recursos genéticos e melhoramento do feijoeiro comum em Pernambuco. In: Queiroz MA, Goedert CO, Ramos SRR, eds. Recursos Genéticos e Melhoramento de Plantas para o Nordeste Brasileiro. Pernambuco, Brasil: Embrapa; 1999. p. 220-230.
- [5] Sudene-Superintendência do Desenvolvimento do Nordeste, Ministério da Integração Nacional. [cited 2017 April 24].
- [6] Marengo JA, Bernasconi M. Regional differences in aridity/drought conditions over Northeast Brazil: present state and future projections. *Climatic Change*. 2014; 129(1):103-115.
- [7] Marengo JA, Torres RR, Alves LM. Drought in Northeast Brazil-past, present, and future. *Theoretical and Applied Climatology*. 2017; 129(3):1189-1200.
- [8] Carbonell SAM. A Cultura do feijão no Brasil: tendências do melhoramento. In: Fancelli AL, Dourado D, eds. Feijão irrigado: estratégias básicas de manejo. São Paulo, Brasil: ESALQ; 1999. p. 9-23.
- [9] MAPA-Ministério da Agricultura, Pecuária e Abastecimento. Descritores. [cited Jun 10 2019]
- [10] Gonçalves DL, Ambrozio VC, Barelli MAA, Neves LG, Sobrinho SP, Luz PB, Silva CR. Divergência genética de acessos tradicionais de feijoeiros através de características da semente. *Bioscience Journal*. 2014; 30(6).
- [11] Mensack MM, Fitzgerald VK, Ryan EP, Lewis MR, Thompson HJ, Brick MA. Evaluation of diversity among common beans (*Phaseolus vulgaris* L.) from two centers of domestication using 'omics' technologies. *BMC Genomics*. 2010; 11:1-11.
- [12] Bitocchi E, Nannia L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, Papa R. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences*. 2012; 109: E788-E796.
- [13] Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Zeuli PS, Gioia T, Logozzo G, Attene G, Nanni L, Papa R. Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist*. 2013; 197(1):300-313.
- [14] Ariani A, Mier y Teran JCB, Gepts P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Molecular Biology and Evolution*. 2017; 35(1):119-131.
- [15] Singh SP, Gutierrez JA, Molina A, Urrea C, Gepts P. Genetic diversity in cultivated common bean: II. Marker-Based analysis of morphological and agronomic traits. *Crop Science*. 1991; 31(1):23-29.
- [16] Khairallah MM, Sears BB, Adams MW. Mitochondrial restriction fragment length polymorphisms in wild *Phaseolus vulgaris* L.: insights on the domestication of the common bean. *Theoretical and Applied Genetics*. 1992; 84(7-8):915-922.
- [17] Gepts P, Kmiecik K, Pereira P, Bliss FA. Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. I. The Americas. *Economic Botany*. 1988; 42(1):73-85.
- [18] Haley SD, Miklas PN, Afanador L, Kelly JD. Random amplified polymorphic DNA (RAPD) marker variability between and within gene pools of common bean. *Journal of the American Society for Horticultural Science*. 1994; 119(1):122-125.
- [19] Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, Gepts P, Nanni L, Papa R, Attene G. Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytologist*. 2016; 209(4):1781-1794.

- [20] Gioia T, Logozzo G, Marzario S, Zeuli PS, Gepts P. Evolution of SSR diversity from wild types to US advanced cultivars in the Andean and Mesoamerican domestications of common bean (*Phaseolus vulgaris*). PLoS One. 2019; 14(1):e0211342.
- [21] Zuiderveen GH, Padder BA, Kamfwa K, Song Q, Kelly JD. Genome-Wide Association Study of Anthracnose Resistance in Andean Beans (*Phaseolus vulgaris*). PLoS One. 2016; 1-17.
- [22] Hurtado-Gonzales OP, Valentini G, Gilio TAS, Martins AM, Song Q, Pastor-Corrales MA. Fine Mapping of Ur-3, a Historically Important Rust Resistance Locus in Common Bean. G3: Genes Genomes Genetic. 2017; 7:557-569.
- [23] Hart JP, Griffiths PD. Genotyping-by-Sequencing enabled mapping and marker development for the potyvirus resistance allele in common bean. The Plant Genome. 2015; 8:1-14.
- [24] Ariani A, Mier y Teran JCB, Gepts P. Genome-wide identification of SNPs and copy number variation in common bean (*Phaseolus vulgaris* L.) using genotyping-by-sequencing (GBS). Molecular Breeding. 2016; 36:1-11.
- [25] Pallotta M, Warner P, Fox R, Kuchel H, Jefferies S, Langridge P. Marker assisted wheat breeding in the southern region of Australia. In Proc. 10th Int. Wheat Genet. Symp., 1–6 Sept. 2003. (Pogna, N. E., ed.). Paestum, Italy: Istituto Sperimentale per la Cerealicoltura, Rome, Italy. 789–791.
- [26] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016; 33(7):1870-1874.
- [27] Fricot E, François O. Application LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution. 2015; 6:925-929.
- [28] Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23:2633-2635.
- [29] Burle ML, Fonseca JR, Kami JA, 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. 2010; 121:801-813.
- [30] Blair MW, Brondani RPV, Díaz LM, Peloso MJD. Diversity and population structure of common bean from Brazil. Crop Science. 2013; 53:1983-1993.
- [31] Pereira PA, Souza CRB. Tipos de faseolina em “raças” de feijão no Brasil. Pesquisa Agropecuária Brasileira. 1992; 27(8):1219-1221.
- [32] Pereira T, Coelho SMM, Bogo A, Guidolin AF, Miquelluti DJ. Diversity in common bean landraces from south Brazil. Acta Botanica Croatica. 2009; 68(1):79-92.
- [33] Beebe S, Skroch PW, Tohme J, Duque MC, Pedraza F, Nienhuis J. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. Crop Science. 2000; 40:264-273.
- [34] Blair MW, Díaz JM, Hidalgo R, Díaz LM, Duque MC. Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). Theoretical and Applied Genetics. 2007; 116:29-43.
- [35] Koinange EMK, Gepts P. Hybrid weakness in wild *Phaseolus vulgaris* L. Journal of Heredity. 1992; 83(2):135-139.