Genet Resour Crop Evol
DOI 10.1007/s10722-015-0325-4

RESEARCH ARTICLE

Genetic diversity of *Capsicum chinense* accessions based on fruit morphological characterization and AFLP markers

Viviane Yumi Baba · Katia Real Rocha · Gisely Paula Gomes · Claudete de Fátima Ruas · Paulo Maurício Ruas · Rosana Rodrigues · Leandro Simões Azeredo Gonçalves

Received: 30 January 2015 / Accepted: 14 September 2015

Abstract *Capsicum chinense* is one of the most cultivated and consumed chili pepper species in Brazil, and Amazon Basin is considered domestication center for the species. *C. chinense* is known for the impressive morphological fruit variability, which can be characterized by different shapes, colors, sizes, and pungency levels. In this study we report the characterization of 71 *C. chinense* accessions from different Brazilian geographic regions using fruit morphological descriptors and AFLP molecular markers. Fourteen descriptors, eight qualitative and six quantitative, were used to fruit characterization. For AFLP analysis, seven combinations of *Eco*RI and *Mse*I primers were tested, and the following combinations were selected: E-ACA/M-CAC, E-ACC/M-CAA, and E-ACG/M-CAA. Morphological data were analyzed using WARD-MLM procedure, while Ward clustering and Bayesian procedure were used for molecular analysis. Variability was found in *C. chinense* in Brazil in terms of fruit phenotype, resulting in three clusters. Fruit shape and fruit weight characteristics were essential for distributing the accessions. Molecular data produced 302 polymorphic bands, forming two groups. It was not possible to group the accessions solely based on their origin using the fruit morphological data and molecular data. There was also no association between the morphological descriptors and AFLP markers. The lack of correlation suggests that both characterization steps are important for understanding and differentiating the *C. chinense* accessions. The combination of morphological and molecular analyses is suggested for the complete and detailed characterization of germplasm databases.

Keywords Chili pepper · Gene bank · Molecular markers · Morphological descriptors · Ward-MLM analysis

V. Y. Baba · G. P. Gomes · L. S. A. Gonçalves (✉)
Agronomy Department, Universidade Estadual de Londrina (UEL), Rodovia Celso Garcia Cid, Km 380, Londrina, PR 86051-900, Brazil
e-mail: leandrosag@uel.br

K. R. Rocha · C. de Fátima Ruas · P. M. Ruas
Biology Department, Universidade Estadual de Londrina (UEL), Rodovia Celso Garcia Cid, Km 380, Londrina, PR 86051-900, Brazil

R. Rodrigues
Plant Breeding Department, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, RJ 28013-602, Brazil

Introduction

Sweet and chili peppers (*Capsicum* spp.) comprise a complex of taxa (species and varieties) from intertropical America, displaying great genetic and phenotypic diversity (Nicolai et al. 2013). There are 32 described taxa, of which five are considered domesticated: *C. annuum* L., *C. baccatum* L. (var. *pendulum*), *C. chinense* Jacq., *C. frutescens* L., and *C. pubescens*

Published online: 29 September 2015

Ruiz et Pav. (DeWitt and Bosland 2009). However, two new wild species, *C. caatingae* Barboza et Agra, and *C. longidentatum* Agra et Barboza (Barbosa et al. 2011) were found in Brazil and probably other species will be identified in the next years.

In Brazil, chili peppers are part of the culture and considered as valuable genetic resources. Chili peppers are mainly grown by small farmers, and an immense variety of sizes, colors, flavors, and pungency levels can be found in their fruits (Sudré et al. 2010). Moreover, Brazil is an important center of diversity for *Capsicum* spp. because it harbors domesticated species as well as semi-domesticated and wild species (Barboza and Bianchetti 2005).

Among the domesticated taxa, *C. chinense* is considered to be the most Brazilian chili pepper because the Amazon Basin is its center of domestication (Pickersgill 1971; Moses and Umaharan 2012; Moses et al. 2014). This species, known for the morphological fruit variability, is the most produced and consumed chili pepper in Brazil (Lannes et al. 2007; Moura et al. 2010; Teodoro et al. 2013). There are many different names applied to chili pepper fruits depending on geographic area in which *C. chinense* is cultivated. In Brazilian conditions, the different names are widely used by farmers, consumers and industry to identify fruit phenotypes. For instance, in a local market in Brazil, depending on fruit phenotype, *C. chinense* can be referred as *pimenta-de-cheiro* (meaning "smelling chili pepper"), *pimenta-de-bode* ("goat chili pepper"), *cumari-do-Pará* ("cumari of Pará state"), *murupi* ("murupi chili pepper"), habanero, and *biquinho* (Carvalho et al. 2014), among other names. However, it is important to point out that these names are not related with any botanical or taxonomical description. Although largely accepted by local markets, this nomenclature should be avoided by the ones who want a more precisely and accurate botanical description, since using shape and fruit color, or even other fruit trait, are not the right manner to identify a *Capsicum* species. Nevertheless, when studying *Capsicum* accessions, it is relevant to include some ethnobotanical aspects especially when it represents strong local inhabitants' self-expressions.

Conservation of diversity in gene bank is extremely important due to high occurrence of genetic erosion caused by destruction of natural habitats (for example, deforestation, desertification, urban expansion, and agricultural modernization), by the replacement of local varieties with genetically improved cultivars and by the abandonment of agricultural activity. Conservation efforts help to create an extensive source of genetic resources available to researchers, providing genes that can enable adaptation to different biotic and abiotic stresses (Ballina-Gómez et al. 2013; Nicolai et al. 2013; Schreinemachers et al. 2014).

To make the accessions conserved in gene bank available for breeding programs, they should be properly described and evaluated (Sudré et al. 2010). Germplasm can be characterized based on morphological descriptors, agronomic traits, and molecular markers (Gonçalves et al. 2008). Among molecular markers, amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995) have numerous advantages. These markers allow obtaining a large number of fragments with high reproducibility, and they are considered an excellent marker for mapping, identifying genotypes, and calculating genetic distance (Albrecht et al. 2012a).

The present study describes the characterization of 71 *C. chinense* accessions from different regions of Brazil using fruit morphological descriptors and AFLP molecular markers. The results of the study add to the knowledge of genetic diversity patterns in *C. chinense* accessions, thus helping to conserve this species and also identify accessions with potential to be used in plant breeding programs.

Materials and methods

Plant material

Seventy-one *C. chinense* accessions from eight Brazilian states representing four geographic regions of Brazil (Fig. 1) from the gene bank of the Universidade Estadual de Londrina (UEL), Londrina, Paraná, Brazil (23°22′S, 51°10′W; altitude of 585 m) were studied.

Phenotyping: fruit descriptors characterization

The accessions were sowed in polystyrene trays with organic plant substrate and plantlets were individually transplanted to plastic pots containing a mixture of soil and substrate (2:1 ratio) after 30 days and were grown in a greenhouse at the UEL.

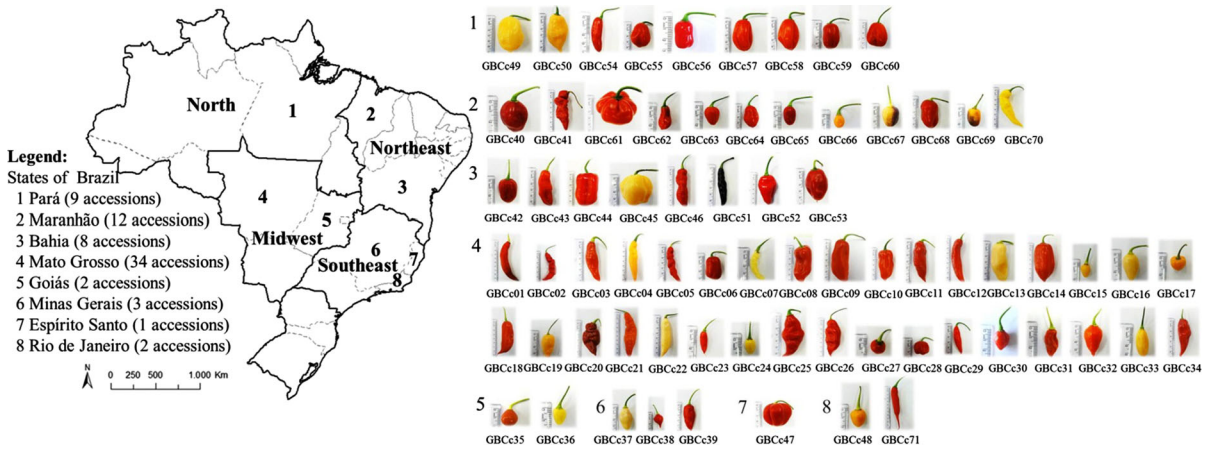

Fig. 1 Geographic distribution of 71 *C. chinense* accessions in Brazil of gene bank of Universidade Estadual de Londrina (UEL)

Fourteen descriptors, eight qualitative and six quantitative, established by the IPGRI (1995), currently Bioversity International, were used to characterize the fruits of the *C. chinense* accessions. Eight qualitative descriptors were observed: fruit color at mature stage; fruit shape; fruit surface; neck at base of fruit; fruit shape at blossom end; fruit blossom end appendage; anthocyanin spots or stripes, and fruit cross-section corrugation. Moreover, six quantitative descriptors were evaluated: fruit length (cm), fruit width (cm), fruit weight (g), fruit placenta (cm), fruit wall thickness (cm), and number of locules.

Genotyping *C. chinense* accessions: AFLP markers

Leaves of plants at the same development stage were sampled. All samples were placed in silica gel and kept in a desiccator for at least 5 days, until the leaves were completely dry. Approximately 150 mg of dry leaves was weighed in a microtube and ground using a MM400 apparatus (Retsch) with a frequency of 30 r/s for 5 min.

DNA was extracted according to the modified protocol by Ferreira and Grattapaglia (1998), using CTAB buffer combined with isopropanol precipitation. All samples were treated with RNase. DNA integrity was confirmed by electrophoresis in a 1 % agarose gel. Concentration and purity were determined by spectrophotometry using a NanoDrop® 2000/2000c (Thermo Fisher Scientific).

The AFLP technique was performed using the protocol described by Vos et al. (1995), with modifications. Approximately 700 ng of DNA from each accession was doubly digested with *Eco*RI and *Mse*I enzymes (5 U each), in the presence of 2 µL of 10X *Mse*I assay buffer, in a final volume of 20 µL, incubated for 18 h at 37 °C. The fragments generated were ligated to linker adaptors for *Eco*RI (0.5 µM) and *Mse*I (5 µM) using the T4 DNA ligase enzyme (1 U); 1X T4 DNA ligase buffer; NaCl (0.05 M); BSA (50 ng/µL); and DTT (0.25 mM) in a final volume of 10 µL. The reaction was incubated in a thermocycler at 37 °C for 3 h, 17 °C for 30 min, and 70 °C for 10 min. The restriction/binding reaction was diluted 1:4 ultrapure water.

The fragments were amplified with a pair of pre-selective primers containing a selective base. Pre-selective amplification was performed in a final volume of 10 µL, using 3.5 µL of the GoTaq® kit *Green Master Mix* (Promega); 0.58 µL of the pre-selective primer (4.75 µM); and 3.0 µL of the dilution of the restriction/binding reaction. The thermocycler program was as follows: 2 min at 72 °C, followed by 20 cycles of 1 s at 94 °C, 30 s at 56 °C, and 2 min at 72 °C, and finally, 30 min at 60 °C. The amplified product was diluted 1:16 in ultrapure water.

A 2.5 µL aliquot of the diluted pre-selective product was used for selective amplification, using 0.54 µL of each selective primer, *Mse*I (5 µM) and *Eco*RI (1 µM); and 3.5 µL GoTaq® *Green Master Mix* (Promega), in a final volume of 10 µL. The selective reactions were performed in a thermocycler under the

following conditions: initial cycle of 2 min at 94 °C; 30 s at 65 °C; and 2 min at 72 °C; 8 cycles of 1 s at 94 °C, 30 s at 64 °C, and 2 min at 72 °C, decreasing 1 °C every cycle; 23 cycles of 1 s at 94 °C, 30 s at 56 °C, and 2 min at 72 °C, and finally, 30 min at 60 °C. Seven combinations of *Eco*RI and *M*seI primers were tested (E-ACA/M-CAC, E-AAC/M-CTA, E-ACC/M-CTA, E-ACC/M-CAA, E-AAG/M-CTA, E-AGG/M-CAC, E-ACG/M-CAA), containing three selective nucleotides, and products were visualized on a 7 % polyacrylamide gel. Of the combinations tested, the three that allowed the greatest number of polymorphic bands were selected (E-ACA/M-CAC, E-ACC/M-CAA, E-ACG/M-CAA). The *Eco*RI primers for these primer sets were labeled with fluorophores and were subjected to capillary electrophoresis in a 3500 xL automated system (Applied Biosystems). The fragment sizes generated were determined by comparing with GeneScan™ 500 ROX® (Applied Biosystems), which was used as a standard for fragments between 35 and 500 bp. The results of electrophoresis of the fragments were combined in a binary matrix using GeneMapper® v.4.1 software (Applied Biosystems). All of the amplifications were performed in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems).

Data analysis

Fruit morphoagronomic descriptors, consisting in qualitative and quantitative variables, were analyzed simultaneously, using the Ward-MLM procedure for clustering the accessions using CLUSTER and IML in the SAS software. A distance matrix was determined by using the Gower algorithm (Gower 1971) to obtain Ward clustering. The ideal number of groups was defined according to the pseudo-F and pseudo-t^2 criteria together with the likelihood profile associated with the likelihood ratio test. According to Mingoti (2007) and Cabral et al. (2010), these parameters (pseudo-F and pseudo-t^2) are similar to a hypothesis test in which each clustering step is related to a test to compare mean vectors of the two clusters jointed to form a new group. Nevertheless, focus should be given to larger values of pseudo-F and pseudo-t^2, since they are related to the least probability of test significance and rejecting the equality of means with major significance. If the equality of mean vectors is rejected, two groups will not be united again in another different group. This approached was also used in gene bank analysis by Gonçalves et al. (2009) and Brasileiro et al. (2013).

For AFLP analysis, the data were interpreted as presence or absence of bands, generating a binary matrix. The Jaccard similarity coefficient was used to estimate the genetic distances between accessions. A representation of the genetic distances between accessions was obtained using the Ward clustering (1000 bootstrapped). The correlation between the molecular (Jaccard) and morphological (Gower) distance measures was found using the Mantel test with 1000 permutations. The analyses were performed using the R statistical computing environment (http://www.r-project.org).

To verify the possible clustering between accessions with molecular data a Bayesian analysis was performed using Structure 2.3.4 software (Pritchard et al. 2000), according to method described by Evano et al. (2005), with burn-in of 10,000 repetitions and Markov Chain Monte Carlo (MCMC) simulations of 100,000 repetitions. K values ranging from 1 to 71 were tested, with 17 independent interactions for each grouping. The determination of probable K clusters was held using Structure Harvester (Earl and vonHoldt 2012).

Results and discussion

Morphoagronomic characterization data

Ten different colors were found for the mature fruit, ranging from white to black, which means that only two colors mentioned on IPGRI descriptors were not found in observed fruits. Red was the most common color (55 %), followed by orange (10 %) and only one accession was observed for white, orange-yellow, and black fruit. Elongated fruit shape was predominant, followed by campanulate, square, triangular, and rounded with 25, 24, 17, 15, and 11 %, respectively. Some fruits had shapes that did not fit any of the shapes proposed in the descriptors list from IPGRI (1995), such as accessions GBCc33, GBCc65, GBCc67, and GBCc69, which had an almost elliptical shape (Fig. 1). Similarly, Sudré et al. (2010), also studying *Capsicum* accessions collected in Brazil, identified two *C. chinense* accessions with elliptical fruit shape that do not correspond to the descriptors proposed by

IPGRI (1995). This result clearly demonstrate the large amount of variability found in *C. chinense* in Brazil in terms of fruit phenotype, since all categories proposed by IPGRI descriptors were observed along with shapes not described on the list.

A semi-rough fruit surface was predominant (49 %), followed by smooth (37 %) and rough (14 %) surfaces. Forty-four percent of the accessions exhibited the neck at base of fruit descriptor. This descriptor is useful to identify *C. chinense* species when comparing with *C. frutescens*. The fruit blossom end appendage was only present in GBCc53 and GBCc64, and anthocyanin spots were found in GBCc19, GBCc37, GBCc67, and GBCc69 accessions. Regarding fruit shape at blossom end, most of the accessions had a truncated shape (46 %), followed by blunt (41 %), sunken (8 %), and sunken and pointed (4 %). For the fruit cross-section corrugation descriptor, the predominant traits were intermediate and corrugated, both representing 42 %, followed by slightly corrugated (15 %).

Large variability was observed for quantitative fruit descriptors mean values. Fruit length ranged from 1.14 to 9.88 cm (mean value 4.20 cm); fruit diameter varied from 0.84 to 3.86 cm (mean value 1.93 cm); fruit weight, with values ranging from 0.46 to 24.18 g (mean = 6.16 g); number of locules from 2 to 4.20 (mean = 3.04); placenta fruit variation from 0.06 to 0.26 cm (mean = 0.13 cm), and wall thickness ranging from 0.10 to 0.36 cm (mean = 0.20 cm).

According to pseudo-F and pseudo-t^2 criteria, the optimum number of groups was three, since maximum values for the tests were achieved at this point. Also, the likelihood profile showed that the highest increase of the likelihood function was also corresponding at the formation of three clusters with an increase of 61.13 (Table 1; Fig. 2), agreeing with the optimum number of groups indicated by pseudo-F and pseudo-t^2 approaches. The likelihood function analysis can define more precise criteria underlying the grouping, resulting in the determination of less subjective groups (Gonçalves et al. 2009). This strategy was also used by Barbé et al. 2010) to determine ideal number of groups when estimating the diversity of 120 recombinant snap bean lines from $F_{6:7}$ generation, finding that the increase in the probability function was greatest when three groups were considered. Also, Ortiz et al. (2008), studying maize landraces, found that increases in the probability function were largest when four and eight groups were formed.

Table 1 Number of groups formed by the Ward-MLM strategy, based on the logarithmic probability function (log-likelihood) and its increment

Number of groups	Log-likelihood	Increase
1	−419.2231	0.0000
2	−384.5947	34.6284
3	−323.4650	61.1297*
4	−296.5926	26.8724
5	−278.8718	17.7209
6	−260.5897	18.2820
7	−259.0785	1.5113
8	−255.0548	4.0237
9	−233.5161	21.5387
10	−211.2340	22.2820

* Greatest increase

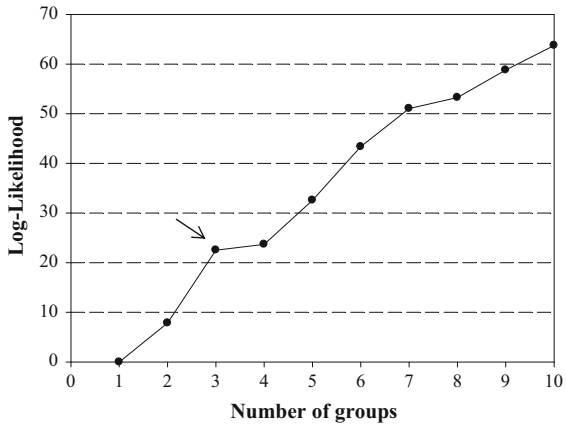

Fig. 2 Graph of the logarithmic function of probability function (log-likelihood) showing the optimum number of groups for 71 *C. chinense* accessions, characterized using morphoagronomic descriptors

The group G1 comprised 16 accessions (GBCc6, GBCc8, GBCc9, GBCc14, GBCc28, GBCc42, GBCc44, GBCc45, GBCc47, GBCc49, GBCc53, GBCc55, GBCc57, GBCc60, GBCc61, and GBCc68), in which the following descriptors were predominant: red fruit color (88 %), square shape (63 %), smooth surface (63 %), truncated fruit tip (50 %), and fruit with wrinkling of the corrugated cross-section (75 %) (Table 2). Only the GBCc53 accession exhibited neck at base of fruit and fruit blossom end appendage. Higher means were found for fruit diameter (2.89 cm), fruit weight (11.86 g), number of locules (3.42),

Table 2 Variables and accessions number per group for categorical traits in each of the three groups (G1, G2 and G3) formed by the Ward-MLM strategy from 71 *C. chinense* accessions

	Accession (71)		
	G1 (16)	G2 (27)	G3 (28)
Fruit color at mature stage			
White	–	–	1
Lemon-yellow	–	2	2
Pale orange-yellow	–	1	–
Orange-yellow	–	2	3
Pale orange	2	2	3
Orange	–	–	3
Light red	–	2	3
Red	14	12	13
Dark red	–	5	–
Purple	–	–	–
Brown	–	–	–
Black	–	1	–
Fruit shape			
Elongate	–	16	2
Almost round	–	–	8
Triangular	4	1	6
Campanulate	2	10	5
Blocky	10	–	2
Others	–	–	5
Fruit surface			
Smooth	10	–	16
Semiwrinkled	6	17	12
Wrinkled	–	10	–
Neck at base of fruit			
Absent	15	3	22
Present	1	24	6
Fruit shape at blossom end			
Pointed	1	22	6
Blunt	8	5	20
Sunken	5	–	1
Sunken and pointed	2	–	1
Fruit blossom end appendage			
Absent	15	27	27
Present	1	0	1
Anthocyanin spots or stripes			
Absent	16	27	24
Present	0	0	4
Fruit cross-sectional corrugation			
Slightly corrugated	–	4	7
Intermediate	4	10	16
Corrugated	12	13	5

placenta fruit (0.20 cm), and wall thickness (0.27 cm) (Table 3).

Group G2 comprised 27 accessions (GBCc1, GBCc2, GBCc3, GBCc4, GBCc5, GBCc7, GBCc10, GBCc11, GBCc12, GBCc13, GBCc18, GBCc20, GBCc21, GBCc22, GBCc25, GBCc26, GBCc29, GBCc34, GBCc39, GBCc41, GBCc43, GBCc46, GBCc50, GBCc51, GBCc52, GBCc70, and GBCc71). Red fruit (44 %), with elongated shape (59 %), semi-rough (63 %) and rough surfaces (37 %), presence of neck at the base of the fruit (89 %), and pointed shape at the fruit tip (81 %) are some of the characteristics of this group (Table 2). The following were relevant traits for the group: longer mean fruit length (6.20 cm) and a greater variety of fruits colors. Eight different color classes were observed in this group.

Group G3 is characterized by clustering fruit with red color (46 %), rounded shape (29 %), smooth surface (57 %), lack of a neck at the base of the fruit (79 %), truncated tip shape (71 %), and wrinkling of the intermediate cross-section (57 %) (Table 2). This group contained the majority of the accessions, as follows: GBCc15, GBCc16, GBCc17, GBCc19, GBCc23, GBCc24, GBCc27, GBCc30, GBCc31,

Table 3 Means for quantitative traits for each of the three groups formed by the Ward-MLM strategy and the first two canonical variables determined from 71 *C. chinense* accessions

Variable	Groups			CAN1	CAN2
	G1	G2	G3		
FL	4.09	9.98	3.84	−0.62	0.69
FWi	0.58	0.55	0.40	0.39	0.78
Fwe	5.39	2.90	1.55	0.23	0.86
NL	0.43	0.47	0.55	0.24	0.40
PT	0.03	0.03	0.02	0.59	0.75
FWT	0.04	0.04	0.06	0.40	0.62

FL fruit length, *FWi* fruit width, *FWe* fruit weight, *NL* number of locules, *PT* placental fruit, *FWT* fruit wall thickness

GBCc32, GBCc33, GBCc35, GBCc36, GBCc37, GBCc38, GBCc40, GBCc48, GBCc54, GBCc56, GBCc58, GBCc59, GBCc62, GBCc63, GBCc64, GBCc65, GBCc66, GBCc67, and GBCc69. This group had greater variety of fruit shapes (nine classes), which were predominantly rounded, followed by triangular and campanulate shapes. Accessions with anthocyanin spots were clustered in this group, which was another trait that differed from the other groups. The accessions belonging to this group had the lowest values for all of the quantitative traits.

Overall, it was not possible to group the accessions solely based on their geographical origin, since accessions from different geographic area were scattered in different clusters. For example, 34 accessions from Mato Grosso State were allocated in three groups (G1, G2, and G3, with 5, 18, and 11 accessions, respectively). The same situation was observed by Finger et al. (2010), which found that geographic distances are uncorrelated with genetic distances among *C. chinense* accessions.

The lack of a relationship between origin and divergence of chili peppers may be due to gene flow caused by natural dispersal agents, such as insects or even wind, responsible for transferring pollen among different plant populations. Although *C. chinense* is considered an autogamous species, a certain amount of cross pollination can be observed in some conditions including high temperatures, wind, and insects' presence. This feature can be a serious problem for seed maintenance in gene banks during seed renovation, since uncontrolled crosses can compromise the genetic purity of the accession, generating unexpected, and in this case undesirable, natural hybrids. Also, it is known that birds can play an important role as *Capsicum* seeds dispersal agent, contributing to the dynamic of changing plant population plants. According to Albrecht et al. (2012b), other very important aspect that might help to explain the lack of correlation among accessions and local origin is the gene flow caused by human activities that transport plants between regions.

In contrast, some morphological descriptors were essential in clustering the accessions, particularly fruit shape. Group G1 had predominantly square fruits, group G2 had elongated fruits, and group G3 had rounded fruits. Fruits with a neck at the base were predominant in G2, and fruits with anthocyanin spots were allocated into group G3. The mean fruit weight variable was higher for group G1 (11.86 g), followed by groups G2 and G3, with 6.88 and 2.21 g, respectively.

Using canonical variable analysis, the following quantitative traits contributed most to explaining the genetic variability in the accessions. The following loadings applied to the first variable: fruit placenta (0.59), wall thickness (0.40), and fruit width (0.39). The following loadings applied to the second variable: fruit weight (0.86), fruit diameter (0.78), and fruit placenta (0.75) (Table 3).

The two first canonical variables explained 92.20 % of the variability among the groups. This value indicates that graphical representation of the first two canonical variables is appropriate for visualizing the relationships between groups and between accessions within the same group (Fig. 3). The graphic representation will allow a satisfactory interpretation of the variability observed among the accessions. Sudré et al. (2010) found that 90.5 % of the total variation was explained by the first two canonical variables in *Capsicum* spp. accessions. Clustering can be observed in the graphical representation of the first two canonical variables. Of the three groups formed, groups G2 and G3 were closer to each other than to G1 (Fig. 2). Similar results can be observed when considering the estimated distances between the groups using the Ward-MLM strategy. The shortest distance observed was between groups G2 and G3, which had a distance of 14.83. The longest distance occurred between groups G1 and G2, which had a distance of 23.03.

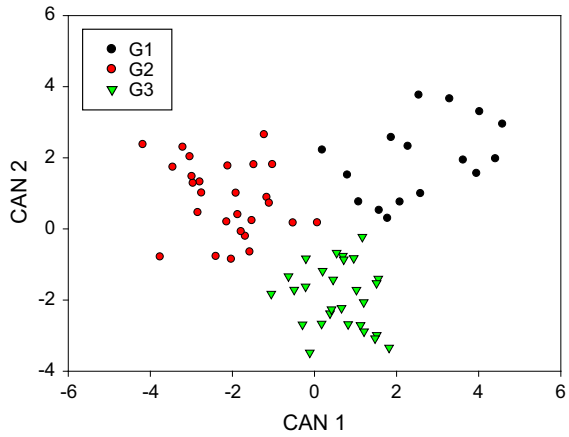

Fig. 3 Dispersion of the first two canonical variables (CAN) with formation of three groups (G1–G3) by the Ward-MLM strategy, considering 71 *C. chinense* accessions

Molecular data

The AFLP markers were efficient in detecting genetic variability among the accessions evaluated in this study. The three combinations of AFLP primers produced 302 polymorphic bands, which were distributed between 60 and 500 bp. The E-ACA/M-CAC, E-ACC/M-CAA, and E-ACG/M-CAA combinations produced 60, 118, and 124 bands, respectively.

Analysis of the frequency distribution of pairwise dissimilarity distances of the 71 *C. chinense* accessions revealed that the distribution was normal, ranging from 0.40 to 1 with a mean distance of 0.73 (±0.10) (Fig. 4). The 0.60–0.70 and 0.70–0.80 classes had the highest frequencies, which were 31.59 and 36.58 %, respectively. The GBCc05 and GBCc65 accessions were the farthest apart (0.99), whereas the GBCc03 and GBCc04 accessions were the closest together (0.43).

Two groups were identified using Ward hierarchical clustering based on the Jaccard genetic distance (Fig. 5). Group I contained the majority of the accessions (46), including the ones from the states of Mato Grosso, Maranhão, Bahia, Pará, Minas Gerais, Goiás and Espírito Santo (25, 7, 6, 4, 2, 1 and 1 from each state, respectively). In Group II were also allocated accessions of different states, including Mato Grosso, Pará, Maranhão, Minas Gerais, Bahia, Rio de Janeiro and Goiás (8, 5, 5, 2, 2, 2, and 1 from

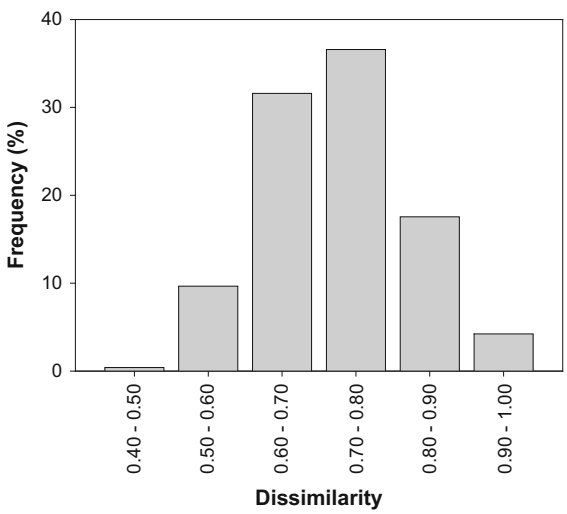

Fig. 4 Frequency distribution of the dissimilarity based on AFLP markers among the 71 *C. chinense* accessions

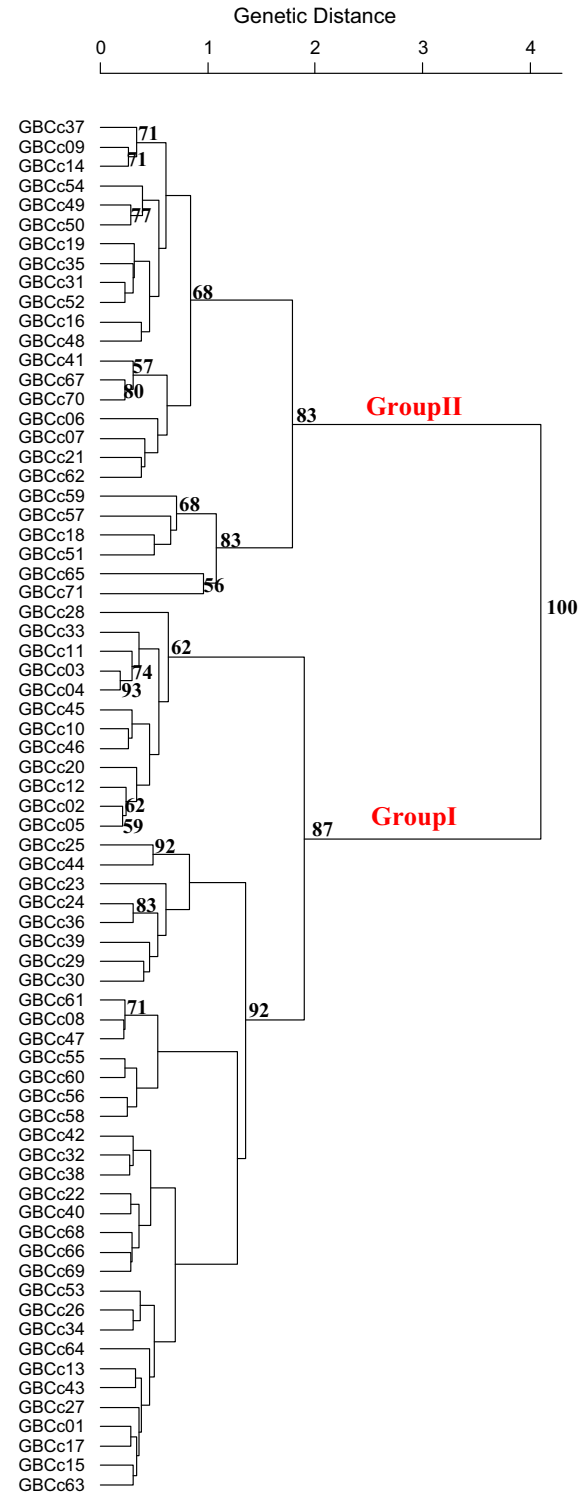

Fig. 5 Dendrogram demonstrating genetic relatedness of 71 *C. chinense* accessions. Cluster analysis conducted using Ward with Jaccard—derived pairwise genetic distances

each state, respectively). Both groups showed a wide variability for the fruit descriptors, and it was not possible to cluster the accessions in relation to fruit phenotypes corresponding to the different shapes and colors that local Brazilian farmers and consumers used to refer which one of the accessions found in Brazil (i.e., *pimenta-de-cheiro*, *pimenta-de-bode*, *cumari-do-Pará*, *murupi*, habanero, and *biquinho*).

From simulations performed by the Structure software using Bayesian procedure, two groups were defined using K clusters approach, according to Evano et al. (2005) (Fig. 6). The groups obtained by Bayesian analysis were partially corroborated by the Ward hierarchical clustering, in which the accessions of group I and II of the Ward clustering were allocated with Bayesian cluster I and II, respectively (Fig. 6, in red and green, respectively), except for the BCc13, GBCc23, GBCc24, GBCc25, GBCc27, GBCc29, GBCc30, GBCc34, GBCc39, GBCc44 and GBCc53 accessions.

The results observed for AFLP data agreed with morphological data, indicating that geographic distances do not correlate with genetic distances among the *C. chinense* accessions evaluated in the present study. We could not observe any correlation between fruit phenotype with geographical location, even inside the groups. Most likely, the seed exchange among farmers, seeds dispersion by birds, and also fruit transportation by chili peppers consumers from different Brazilian regions are playing an important role in this mixing status in chili pepper around many regions in Brazil. Moses et al. (2014) investigated the genetic diversity of *C. chinense* accessions across a broad geographic region, including accessions from South America and the Caribbean islands, using microsatellite markers. These authors concluded that cluster analysis resulted in two distinct genetic clusters, corresponding to Upper and Lower Amazon regions, suggesting two independent domestication events or two putative centers of diversity in these regions. In our study, most of the accessions were sampled from Lower Amazon region (only one collected in the Amazonian region, more precisely in state of Pará), including areas with strong anthropic influence in agrosystems, like Mato Grosso and Goiás. Another important observation is regarding the structure of the population sampled in paper published by Moses et al. (2014), in which 22 *C. chinense* accessions from Brazil were sampled and all of them came from AVRDC and USDA, meaning that probably these accessions were collected long time ago and are being kept in environmental conditions quite different from natural conditions in Brazil.

The morphological descriptors and AFLP marker data revealed patterns of distinct dissimilarity among the matrices generated (correlation value = 0.03, non-

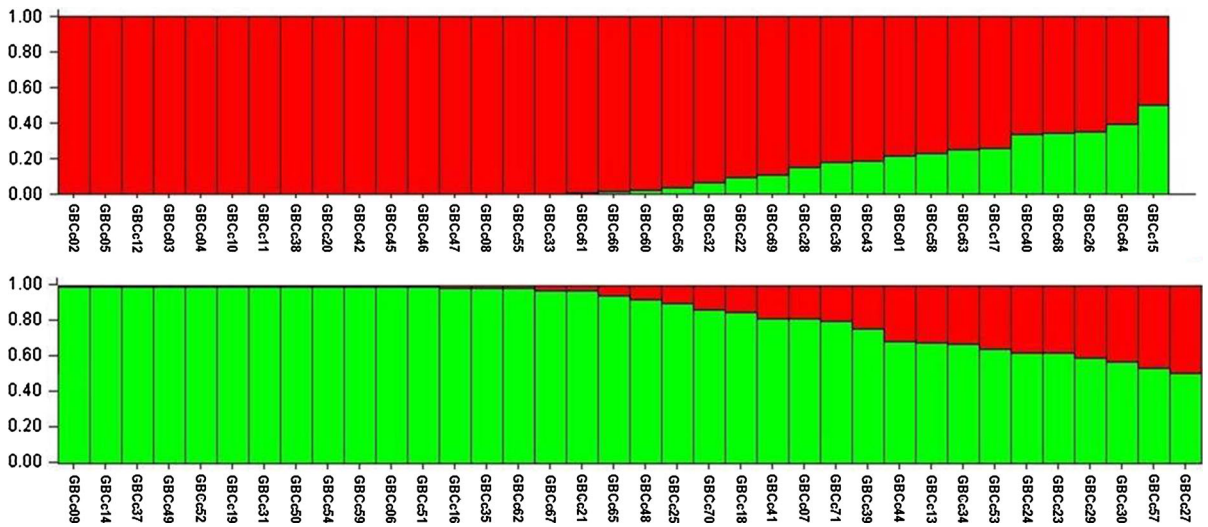

Fig. 6 Assignment of 71 *C. chinense* accessions by the *structure bar plots* based on three AFLP combinations of primer/enzyme (E-ACA/M-CAC, E-ACC/M-CAA, and E-ACG/M-CAA). Two colors represent different clusters. The *y*-axis displays the estimated percentage membership of each accession in a determined cluster

significant). Whereas AFLP markers cover a higher proportion of the genome, including coding and non-coding regions, the phenotypic markers exclusively represent the coding regions. The molecular profile may represent regions uncorrelated with the studied phenotypic traits. The lack of correlation between morphological descriptor and AFLP markers suggests that both characterization steps are important for understanding and differentiating between the *C. chinense* accessions.

References

Albrecht E, Zhang D, Saftner RA, Stommel JR (2012a) Genetic diversity and population structure of *Capsicum baccatum* genetic resources. Genet Resour Crop Evol 59:517–538

Albrecht E, Zhang D, Mays AD, Saftner RA, Stommel JR (2012b) Genetic diversity in *Capsicum baccatum* is significantly influenced by its ecogeographical distribution. BMC Genet 13:68

Ballina-Gómez H, Latournerie-Moreno L, Ruiz-Sánchez E, Pérez-Gutiérrez A, Rosado-Lugo G (2013) Morphological characterization of *Capsicum annuum* L. accessions from southern Mexico and their response to the *Bemisia tabaci–Begomovirus* complex. Chil J Agric Res 73:329–338

Barbé TC, Amaral Júnior AT, Gonçalves LSA, Rodrigues R, Scapim CA (2010) Association between advanced generations in inbred lines of snap bean by the Ward-Modified Location Model. Euphytica 173:337–343

Barbosa GE, Agra MF, Romero MV, Scaldaferro MA, Moscone EA (2011) New endemic species of *Capsicum* (Solanaceae) from the Brazilian Caatinga: comparison with the recircumscribed *C. parvifolium*. Syst Bot 36:768–781

Barboza GE, Bianchetti LB (2005) Three new species of *Capsicum* (Solanaceae) and a key to the wild species from Brazil. Syst Bot 30:863–871

Brasileiro BP, Marinho CD, Costa PMA, Moreira EFA, Peternelli LA, Barbosa MHP (2013) Genetic diversity in sugarcane varieties in Brazil based on the Ward-Modified Location Model clustering strategy. Genet Mol Res 13:1650–1660

Cabral PDS, Soares TCB, Gonçalves LSA, Amaral Júnior AT, Lima ABP, Rodrigues R, Matta FP (2010) Quantification of the diversity among common bean accessions using Ward-MLM strategy. Pesqui Agropecu Bras 45:1124–1132

Carvalho SIC, Ragassi CF, Bianchetti LB, Reifschneider FJB, Buso GSC, Faleiro FG (2014) Morphological and genetic relationships between wild and domesticated forms of peppers (*Capsicum frutescens* L. and *C. chinense* Jacquin). Genet Mol Res 13:7447–7464

DeWitt D, Bosland PW (2009) The complete chile pepper book: a gardener's guide to choosing, growing, preserving and cooking, 1st edn. Timber Press, London

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evano G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análise genética. Embrapa Cenargen, Brasília, p 220

Finger FL, Lannes SD, Schuelter AR, Doege J, Comerlato AP, Gonçalves LSA, Ferreira FRA, Clovis LR, Scapim CA (2010) Genetic diversity of *Capsicum chinensis* (Solanaceae) accessions based on molecular markers and morphological and agronomic traits. Genet Mol Res 9:1852–1864

Gonçalves LSA, Rodrigues R, Amaral Júnior AT, Karasawa M, Sudré CP (2008) Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions. Genet Mol Res 7:1289–1297

Gonçalves LSA, Rodrigues R, Amaral Júnior AT, Karasawa M, Sudré CP (2009) Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using Ward-modified location model. Genet Mol Res 8:364–374

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857–871

IPGRI - International Plant Genetic Resources Institute (1995) Descriptores para *Capsicum* (*Capsicum* spp.). IPGRI, Roma

Lannes SD, Finger FL, Schuelter DR, Casali VWD (2007) Growth and quality of Brazilian accessions of *Capsicum chinense* fruits. Sci Hortic 112:266–270

Mingoti SA (2007) Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. UFMG, Belo Horizonte, p 297p

Moses M, Umaharan P (2012) Genetic structure and phylogenetic relationships of *Capsicum chinense*. J Am Soc Hortic Sci 137:250–262

Moses M, Umaharan P, Dayanandan S (2014) Microsatellite based analysis of the genetic structure and diversity of *Capsicum chinense* in the Neotropics. Genet Resour Crop Evol 61:741–755

Moura MCCL, Gonçalves LSA, Sudré CP, Rodrigues R, Amaral Júnior AT, Pereira TNS (2010) Algoritmo de Gower na estimativa da divergência genética em germoplasma de pimenta. Hortic Bras 28:155–161

Nicolai M, Cantet M, Lefebvre V, Sage-Palloix AM, Palloix A (2013) Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. Genet Resour Crop Evol 60:2375–2390

Ortiz R, Crossa J, Franco J, Sevilla R, Burgueño J (2008) Classification of Peruvian highland maize races using plant traits. Genet Resour Crop Ev 55:151–162

Pickersgill B (1971) Relationships between weedy and cultivated forms in some species of chili peppers (genus *Capsicum*). Evolution 25:683–691

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

Schreinemachers P, Ebert AW, Wu MH (2014) Costing the ex situ conservation of plant genetic resources at AVRDC—The World Vegetable Center. Genet Resour Crop Evol 61:757–773

Sudré CP, Gonçalves LSA, Rodrigues R, do Amaral Júnior AT, Riva-Souza EM, Bento CS (2010) Genetic variability in domesticated *Capsicum* spp as assessed by morphological and agronomic data in mixed statistical analysis. Genet Mol Res 9:283–294

Teodoro AFP, Alves RBN, Ribeiro LB, Reis K, Reifschneider FJB, Fonseca MEN, Silva JP, Agostini-Costa T (2013) Vitamin C content in Habanero pepper accessions (*Capsicum chinense*). Hortic Bras 31:59–62

Vos P, Hogers R, Bleeker M, Reijans M, de Lee TV, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 21:4407–4414