Genet Resour Crop Evol (2017) 64:1775–1788
DOI 10.1007/s10722-017-0535-z

NOTES ON NEGLECTED AND UNDERUTILIZED CROPS

Genetic diversity and structure in a major Brazilian annatto (*Bixa orellana*) germplasm bank revealed by microsatellites and phytochemical compounds

Gabriel Dequigiovani · Santiago Linorio Ferreyra Ramos · Alessandro Alves-Pereira · Eliane Gomes Fabri · Paulo Roberto Nogueira Carvalho · Marta Gomes da Silva · Maria Teresa Vilela Nogueira Abdo · Antônio Lucio Mello Martins · Charles Roland Clement · Elizabeth Ann Veasey

Received: 7 December 2016 / Accepted: 6 June 2017 / Published online: 20 June 2017

Abstract Annatto (*Bixa orellana* L.) is a tropical crop indigenous to the Americas, probably Amazonia. Annatto is commercially valuable in the food and cosmetics industries as a natural dye used instead of synthetic dyes. In addition, annatto contains other important substances for human health, such as geranylgeraniol, tocotrienols and other carotenoids. The aim of the present study was to evaluate the genetic diversity of 63 accessions from the annatto germplasm bank at the Agronomic Institute (IAC), São Paulo, Brazil, using four phytochemical compounds and 16 microsatellite markers. Significant variation was observed for the phytochemical compounds, ranging from 2 to 7.31 g (100 g dry matter—DM)$^{-1}$ for bixin, 2.14–7.11 g (100 g DM)$^{-1}$ for lipids, 0.25–1.05 g (100 g DM)$^{-1}$ for tocotrienols, and 0.49–2.61 g (100 g DM)$^{-1}$ for geranylgeraniol content. A total of 73 alleles was observed in the molecular characterization with 16 microsatellite loci. We found higher expected heterozygosity than observed heterozygosity for all loci, indicating strong deficits of heterozygotes. For both molecular and phytochemical compounds, cluster and PCoA

Electronic supplementary material The online version of this article (doi:10.1007/s10722-017-0535-z) contains supplementary material, which is available to authorized users.

G. Dequigiovani (✉) · A. Alves-Pereira · E. A. Veasey (✉)
Department of Genetics, Luiz de Queiroz College of Agriculture, University of São Paulo, Av. Pádua Dias 11, CP 83, Piracicaba, São Paulo 13400-970, Brazil
e-mail: gabriel.dequi@gmail.com

E. A. Veasey
e-mail: eaveasey@usp.br

S. L. F. Ramos
Instituto de Ciências Exatas e Tecnologia em Itacoatiara, Universidade Federal do Amazonas, Rua Nossa Senhora do Rosário, 3863, Itacoatiara, Amazonas 69103-128, Brazil

E. G. Fabri
Centro de Horticultura, Instituto Agronômico de Campinas, Av. Barão de Itapura, 1481, Campinas, São Paulo 13020-902, Brazil

P. R. N. Carvalho · M. G. da Silva
Instituto de Tecnologia de Alimentos, Av. Cônego Antônio Roccato, 2880, Campinas, São Paulo 13073-001, Brazil

M. T. V. N. Abdo · A. L. M. Martins
Agência Paulista de Tecnologia dos Agronegócios, Pólo Apta Centro Norte, PO Box 24, Pindorama, São Paulo 15830-000, Brazil

C. R. Clement
Instituto Nacional de Pesquisas da Amazônia, Cx. Postal 2223, Manaus, Amazonas 69080-971, Brazil

analyses tended to separate the accessions from Rondônia, northern Brazil, with only a few exceptions, from the Southwestern accessions. The same two groups were found in the Bayesian analysis with molecular data. Rondônia accessions showed higher values for all the phytochemical compounds and higher levels of genetic diversity. Some accessions presented bixin levels well above the average and are promising materials to be used in genetic improvement programs.

Keywords Genetic diversity · SSR markers · Bixin · *Bixa orellana* · Geranylgeraniol · Tocotrienol

Introduction

Annatto (*Bixa orellana* L.) is a crop native to continental tropical America (Arce 1999), probably Amazonia (Sandy-Cuen and Becerra 2003; Clement et al. 2010), with the wild variety (*B. orellana* var. *urucurana*) naturally occurring in open forests always associated with riparian environments, presenting a wide geographical distribution in Amazonia, northern South America and Central America. The domesticated types have been cultivated since ancient times and grow from Mexico to Argentina and throughout the Caribbean Islands in the Americas. The plant is also naturalized and cultivated in tropical and subtropical areas of other continents (de Franco et al. 2008; Leal and de Clavijo 2012; Moreira et al. 2015).

The only recent systematic treatment of *Bixa* (Baer 1976) accepted *B. orellana* L. Sp. Pl. (1753) 512 as the correct specific name, reducing a number of previous names to synonymy: *B. americana* Poir. Encycl. 6 (1804) 229; *B. orleana* Noronha. Verh. Batav. Genootsch. Kunsten 5(4) (1790) 8; *B. katangensis* Delpierre. Taxon 19 (1970) 304; *B. odorata* Ruiz et Pav. ex G. Don. Gen. Hist. 1 (1831) 295; *B. orellana* var. *leiocarpa* (Kuntze) Standl. et L.O.Williams. Fieldiana Bot. 29 (1961) 358; *B. orellana* f. *leiocarpa* (Kuntze) J.F.Macbr. Publ. Field Mus. Nat. Hist., Bot. Ser. 14(4) (1941) 11; *B. purpurea* Sweet. Hort. Brit. 33 (1826); *B. tinctaria* Salisb. Prodr. Stirp. Chap. Allerton (1796) 369; *B. upatensis* Ram. Goyena. Fl. Nicarag. 1 (1909) 217; *Orellana americana* (Poir.) Kuntze. Revis. Gen. Pl. 1 (1891) 44; *O. americana* var. *leiocarpa* Kuntze. Revis. Gen. Pl. 1 (1891) 45; *O. orellana* (L.) Kuntze. Revis. Gen. Pl. 3(2) (1898) 9. This treatment is accepted by the Flora of Brazil (Forzza et al. 2010).

Bixa orellana is a small tree or shrub measuring from 3 to 8 m in height. It presents simple and glabrous (when adult) leaves, measuring on average 8 cm in length and 4 cm in width, with long petioles, arranged alternately along the branches (de Franco et al. 2008; Leal and Clavijo 2010, 2012). The flowers are large, with colors varying from white to several shades of pink and lilac. Flowers are hermaphrodite, with five sepals, appearing at the end of the branches, forming fascicles. The fruit is a dehiscent ovoid capsule, with two or three carpels, covered with flexible spines when juvenile. It may be reddish, greenish or brown, measuring three to five centimeters long, containing many black seeds covered by a reddish waxy aril with characteristic odor. The fruits are arranged in bunches with up to 17 units. A well-developed fruit can contain 40–60 seeds.

Brazil is the major producer of annatto, followed by Peru and Kenya, but the crop is also found in the Dominican Republic, Colombia, Jamaica, Costa Rica, Suriname and other countries in Asia (Akshatha et al. 2011). In Brazil, annatto is grown in several regions, with the states of São Paulo, Rondônia and Pará the largest producers. The largest companies that process annatto are located mainly around the city of São Paulo and in the metropolitan region of Campinas in the state of São Paulo (Fabri 2015). Annatto is commercially valuable due to the application of its pigments in the food and cosmetics industries. The principal pigment in annatto extract is bixin, which is found in the resinous coating of the seed itself (Nisar et al. 2015). This extract is noteworthy because of its lack of toxicity, its intense coloring capacity and its range of colors, comprising red, orange and yellow hues (Alves et al. 2006). Recently, this crop has acquired further importance for containing other important substances for human health, such as geranylgeraniol, tocotrienols and other carotenoids (Albuquerque and Meireles 2012), including the potential treatment of the important tropical and subtropical disease leishmaniasis, caused by *Leishmania braziliensis* and *L. amazonensis*, using annatto's essential oil (Monzote et al. 2006; Lopes et al. 2012). Annatto seed extracts also showed insect repellent properties, including protection against *Aedes aegypti*, the insect vector of yellow fever,

dengue, chikungunya, zika and other diseases (Giorgi et al. 2013).

The use of annatto as a colorant is not new. The Aztecs used annatto extract as a dye for textiles, body paint (such as in lipsticks), and as a food colorant in the drink *cacahuatl* (Giuliano et al. 2003). The annatto seed is characterized by the presence of an aril on its surface that contains multiple substances besides the characteristic red pigment. This aril represents about 5–10% of the seed weight, of which 30% is the carotenoid bixin. The remaining 70% is composed of carbohydrates (32%), lipids (30%), humidity (3.5%), protein (2.5%) and ash (2.0%) (Carvalho et al. 1991). Geranylgeraniol is a naturally occurring linear diterpene soluble in organic solvents, such as chloroform, acetone and alcohol. Geranylgeraniol is an important intermediate of vitamin K, tocopherols and many hormones, and in carotenoid biosynthesis. The presence of geranylgeraniol in annatto seeds was initially described by Craveiro et al. (1989), followed by Jondiko and Pattenden (1989), which established the concentration of approximately 1 g $(100\ g)^{-1}$ of this metabolite in annatto seeds. Tocotrienols are substances that exhibit strong antioxidant activity and are commonly known as vitamin E. According to Tan and Foley (2000), annatto is one of the few plants containing tocotrienols in a much higher proportion than tocopherols, generating great interest from the pharmaceutical industry. Although found in low amounts in the human diet, tocotrienols are abundant in rice (*Oryza sativa* L.), palm oil (*Elaeis guineensis* Jacq.), and annatto. Tocotrienols are neuro-protective, anti-cancer and cholesterol lowering (Sen et al. 2007). Frega et al. (1998) described the presence of tocotrienols in annatto seeds at concentrations of 0.14 g $(100\ g)^{-1}$. The concentration of carotenoids in annatto seeds usually ranges from 3.12 $(100\ g\ DM)^{-1}$ to 6.26 g $(100\ g\ DM)^{-1}$. The most important carotenoid in annatto seed is cis-bixin, a monomethyl ester of dicarboxylic acid corresponding to more than 80% of the total carotenoid content of annatto seeds (Carvalho et al. 1993; Auttachoat et al. 2011). Despite its growing economic importance, the conservation and manipulation of annatto genetic resources represents the main and most difficult goal for this crop. Considered as the center of origin of annatto, Brazil hosts the greatest diversity of this species. Therefore, characterizing this diversity is a priority in order to promote the conservation of genotypes, as well as to provide information for breeding programs (Rodrigues 1995).

Microsatellite markers or simple sequence repeats (SSR) show high polymorphism, co-dominance and multiallelism. Besides presenting highly reproducible results, this marker is widely distributed in the nuclear genome, making it a useful tool for assessing the genetic diversity and structure of plant populations (Kalia et al. 2011; Vieira et al. 2016). Microsatellite primers were recently developed by Dequigiovanni et al. (2014) and this is the first report using this marker to analyze the genetic diversity of annatto accessions in Brazil. Although annatto is an important source of natural dye, there is almost no research addressing the genetic variability within this species and investigations on how this variability is distributed in Brazil. In this context, the aim of this study was to characterize the genetic diversity and structure of accessions from the annatto germplasm bank at the Agronomic Institute (IAC), Campinas, São Paulo, Brazil, using microsatellite markers and four phytochemical compounds.

Materials and methods

Plant material

Sixty-three accessions from the annatto Germplasm Bank of IAC, maintained at the Polo Regional Centro Norte in Pindorama, São Paulo, were evaluated in this study. The accessions originated from the Brazilian Central–West, Southeast, and North regions, as well as one accession from Peru and 20 accessions of unknown origin (Table 1; Fig. 1). The collection is maintained in the field, with plots of six half-sib plants, and young leaves from one plant from each accession were collected for the molecular analysis, conducted at the Genetics Department of Luiz de Queiroz College of Agriculture, University of São Paulo.

Phytochemical analysis

For the phytochemical analysis, annatto seeds from the harvest of 2011 were used. Fruits from the selected plants were harvested and dried in the sun. After drying, the seeds were separated from the fruits (capsules) manually. From 100 to 200 g seeds were placed in plastic pots, which were identified and sent to the laboratory at the Institute of Food Technology

Table 1 Mean values of moisture, lipids, bixin, tocotrienols and geranylgeraniol [g (100 g DM)$^{-1}$] in annatto (*Bixa orellana*) seeds of 63 accessions from the Instituto Agronômico germplasm bank in Campinas, São Paulo, Brazil

ID	Origin	Moisture	Lipids	Bixin	Tocotrienols	Geranylgeraniol
1	Cuiabá-MT	12.94	2.24	3.05	0.41	0.69
2	Cuiabá-MT	8.80	2.87	3.52	0.76	1.13
3	Vale do Ribeira-SP	8.75	2.85	3.57	0.77	1.08
4	Viçosa-MG	10.44	2.63	3.88	0.49	0.99
5	Rezende-RJ	11.36	3.02	2.82	0.80	0.82
6	Sorocaba-SP	9.69	3.43	2.89	0.65	0.96
7	Igaratá-SP	9.96	3.08	2.71	0.39	1.13
8	Tupi Paulista-SP	10.00	3.24	2.60	0.64	1.78
9	Unknown	10.37	3.40	2.75	0.38	1.50
10	Guaraci-SP	8.91	2.97	3.76	0.61	0.98
11	São José do Rio Preto-SP	9.71	3.43	2.82	0.50	1.19
12	Unknown	10.34	3.41	3.18	0.54	1.18
13	Unknown	11.00	3.45	2.46	0.78	1.70
14	Icem-SP	10.14	2.94	3.22	0.74	1.12
15	Unknown	11.27	3.43	2.67	0.59	1.26
16	Minas Gerais-MG	10.88	3.00	4.05	0.48	1.49
17	Barretos-SP	9.23	3.41	2.80	0.51	1.70
18	Unknown	10.61	3.45	2.01	0.46	1.57
19	Unknown	5.78	2.81	2.77	0.43	1.48
20	Icem-SP	9.82	3.38	2.90	0.73	0.93
21	Olímpia-SP	8.73	2.99	2.28	0.65	0.85
22	Peru	11.40	2.48	2.15	0.25	0.56
23	Unknown	12.37	4.66	4.78	1.05	1.10
24	Unknown	9.78	5.00	7.31	1.02	1.57
25	Unknown	10.74	2.55	3.10	0.61	0.63
26	Unknown	10.71	3.36	3.79	0.92	0.82
27	Unknown	8.39	2.14	2.73	0.82	0.49
28	Unknown	10.51	3.42	4.25	0.72	0.77
29	Unknown	8.57	3.28	2.89	0.59	0.88
30	Unknown	11.43	3.16	4.68	0.70	0.76
31	Unknown	10.80	4.40	4.49	1.00	0.62
32	Unknown	11.88	4.01	4.71	0.55	1.21
33	S. João do Pau D'alho-SP	9.60	3.37	3.31	0.93	0.78
34	Unknown	10.11	3.53	3.13	0.43	1.28
35	Unknown	9.46	2.64	3.25	0.62	1.05
36	Unknown	10.61	3.33	2.00	0.47	1.03
37	Unknown	10.31	3.40	3.88	0.82	0.99
38	Corumbiara-RO	12.54	6.91	6.63	1.04	1.82
39	Corumbiara-RO	10.25	4.38	6.56	1.03	1.63
40	Colorado do Oeste-RO	12.67	6.50	4.57	0.94	1.53
41	Colorado do Oeste-RO	12.54	6.43	4.43	0.97	1.78
42	Colorado do Oeste-RO	12.66	3.29	4.02	0.71	1.15
43	Corumbiara-RO	11.32	4.47	4.32	0.63	0.82
44	Corumbiara-RO	12.37	5.53	2.56	0.74	2.45

Table 1 continued

ID	Origin	Moisture	Lipids	Bixin	Tocotrienols	Geranylgeraniol
45	Corumbiara-RO	14.32	4.68	3.35	0.50	1.51
46	Corumbiara-RO	11.33	4.41	3.85	0.65	1.91
47	Corumbiara-RO	10.12	4.29	4.54	1.05	1.46
48	Colorado do Oeste-RO	12.59	4.15	3.39	0.64	0.88
49	Colorado do Oeste-RO	12.27	3.88	4.66	0.66	1.40
50	Colorado do Oeste-RO	9.79	3.11	4.39	0.70	1.11
51	Colorado do Oeste-RO	10.32	3.81	3.67	0.65	1.98
52	Corumbiara-RO	9.92	2.97	4.61	0.66	0.97
53	Colorado do Oeste-RO	8.52	3.39	4.91	0.42	1.30
54	Colorado do Oeste-RO	10.95	3.62	2.77	0.54	1.06
55	Colorado do Oeste-RO	10.87	3.51	4.66	0.61	1.45
56	Colorado do Oeste-RO	12.71	4.90	3.41	0.66	2.61
57	Corumbiara-RO	12.75	4.92	2.68	0.52	2.15
58	Colorado do Oeste-RO	13.14	4.06	4.00	0.97	1.24
59	Colorado do Oeste-RO	11.42	3.17	4.05	0.64	1.49
60	Corumbiara-RO	12.10	5.67	3.99	0.70	1.52
61	Colorado do Oeste-RO	13.98	6.97	5.08	0.64	1.94
62	Colorado do Oeste-RO	12.79	7.11	5.13	0.65	2.04
63	Colorado do Oeste-RO	12.34	2.61	3.08	0.38	0.84
	Overall mean	10.83	3.79	3.69	0.67	1.27

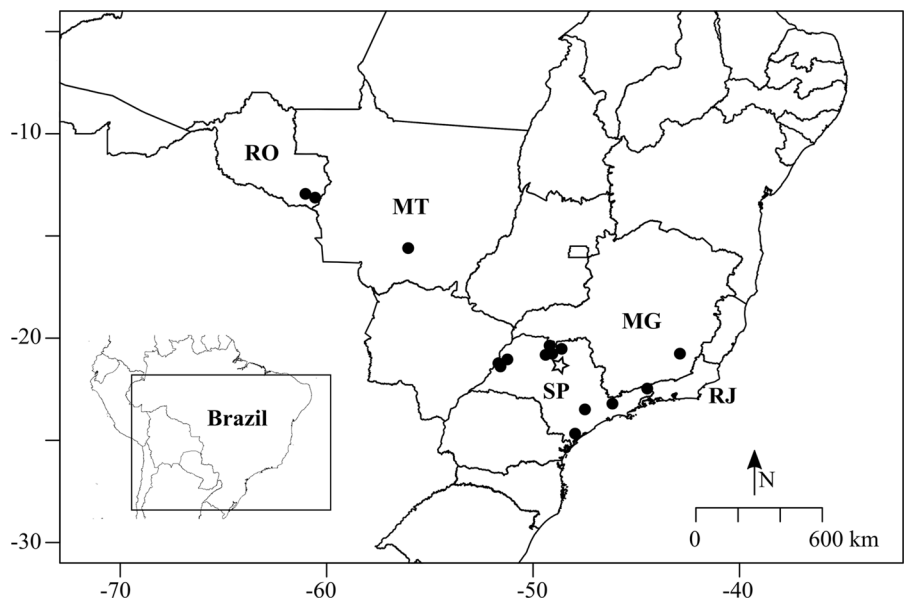

Fig. 1 Map of Brazil indicating the origins of the annatto (*Bixa orellana*) accessions maintained in the IAC Germplasm Bank. *Black circles* represent the origin of accessions, while the *star* indicates the location of the Germplasm Bank in Pindorama, SP. *MG* Minas Gerais, *MT* Mato Grosso, *RJ* Rio de Janeiro, *RO* Rondônia and *SP* São Paulo

(ITAL), Campinas, SP, where the analyses were carried out. In the laboratory, the seeds were then transferred to glass containers, where they were kept away from light and refrigerated until the analyses began.

Moisture determination was based on the method described by AOAC (Horwitz 2005). The determination of lipids was conducted based on the method 2006.06 described by AOAC (Horwitz 2005) using hexane. The analytical method for the determination

of total carotenoids expressed as bixin was based on the saponification of bixin, dilution with potassium hydroxide solution and spectrophotometric quantification, as described by Carvalho et al. (2010).

The analytical method used for the determination of tocotrienols and geranylgeraniol was based on saponification with potassium hydroxide solution, extraction of the unsaponifiable fraction with ethyl ether, and transfer of analytes to n-hexane. The analytical method for the determination of tocotrienols was based on the methodology described by Panfili et al. (2003). The tocotrienols analysis was performed in a Prominence LC-20A liquid chromatograph coupled to a fluorescence detector RF-10AXL (Shimadzu, Tokyo), using the excitation wavelength of 292 nm and emission of 326 nm. The analytes were resolved on a normal phase column of LiChrospher Si60 (12.5 cm long × 4 mm d.i. and 5 µm particle diameter; Merck, Darmstadt, Germany), having as the mobile phase n-hexane: ethyl acetate: acetic acid (97.6:1.8:0.6, v/v/v), in an isocratic system. The determination of geranylgeraniol was based on the methodology described by Zahn et al. (2000). An Infinity 1260 liquid chromatograph and diode array detector (Agilent, USA), with monitoring at 210 nm, were used for the geranylgeraniol determination. A LiChrospher 100RP-18 column (12.5 cm long, 4 mm d. i. and 5 µm particle diameter; Merck, Darmstadt) was used, and the mobile phase was composed of methanol: 20 mM ammonium acetate (90: 10, v/v), in an isocratic system.

Molecular analysis

DNA was extracted from recently expanded young leaves according to Doyle and Doyle (1990). DNA was quantified by comparison with known concentrations of standard DNA (lambda DNA; Invitrogen, Carlsbad, CA, USA) in electrophoresis agarose gels (1%) stained with GelRed (Biotium, Fremont, CA, USA).

Sixteen SSR markers developed for *B. orellana* (BorA2, BorA3, BorA5, BorB1, BorB4, BorB5, BorB12, BorC5, BorD1, BorD2, BorF9, BorG4, BorG11, BorH3, BorH7, BorH10) (Dequigiovanni et al. 2014, submitted) were used in the present study. These markers were selected based on their polymorphism and compatibility for multiplexing. An M13 sequence tail was added to the 5′ end of each forward primer following a labeling protocol (Schuelke 2000). Microsatellite fragments were amplified using a MyCycler Thermal Cycler (Bio-Rad, Hercules, CA, USA) in a total reaction volume of 10 µL, containing 20 ng of genomic DNA template, 1 U *Taq* DNA polymerase (Fermentas, Vilnius, Lithuania), 1X polymerase chain reaction buffer (10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.08% Nonidet P40), 0.25 mM each dNTP, 1.5 mM $MgCl_2$, 2.5 pmol of forward and M13 label primers (FAM, HEX or NED dyes) and 5 pmol of reverse primers.

Polymerase chain reactions were carried out according to Schuelke (2000), consisting of 94 °C (5 min), then 30 cycles at 94 °C (30 s)/T_a °C (45 s) (T_a = annealing temperature)/72 °C (45 s), followed by 8 cycles at 94 °C (30 s)/53 °C (45 s)/72 °C (45 s), and a final extension at 72 °C for 10 min. Quality of amplification was checked by electrophoresis in agarose gels (1.5%) stained with GelRed (Biotium). Fragment separation and detection were performed on an ABI Prism 3130xl capillary sequencer using GeneScan 500 Rox-labelled size standard (Applied Biosystems, Foster City, CA, USA). SSR patterns were scored using the Genemapper v4.0 software (Applied Biosystems).

Statistical analyses

Phytochemical diversity analyses

The dispersion of phenotypic diversity among accessions was evaluated by Principal Components Analysis (PCA), carried out with the R Statistical Software–ade4 package (Dray and Dufour 2007). Cluster analysis was performed using Euclidean distances, and the UPGMA (Unweighted pair group method with arithmetic mean) clustering method, implemented in R Statistical Software–stats package (R Core Team 2015) and formatted with FigTree (Rambaut and Drummond 2010).

Genetic diversity and structure analyses

Genetic diversity parameters, including total number of alleles (A), allelic richness (Ar), observed (H_O) and expected (H_E) heterozygosities, number of private alleles (Ap) and inbreeding coefficient (F_{IS}), were estimated for each locus using diveRsity package from the R project (Keenan et al. 2013). Genetic distances

between individuals were estimated by Rogers' genetic distance (Rogers 1972). The resulting distance matrix was used to construct a dendrogram with the Neighbour-joining algorithm (Saitou and Nei 1987), with 1000 bootstrap replicates, implemented in Population 1.2.32 (Langella 2002). The final dendrogram was formatted with FigTree (Rambaut and Drummond 2010). The dispersion of genetic diversity among accessions was evaluated by Principal Coordinate Analysis (PCoA), carried out with the R Statistical Software–ade4 package (Dray and Dufour 2007). The apparent outcrossing rate ($\hat{t}_a$) was estimated by using the estimated inbreeding coefficient according to the equation $\hat{t}_a = (1-f)/(1+f)$ (Weir 1996).

We also analyzed the population structure and detection of admixture using a Bayesian model based on the clustering method implemented in Structure 2.3.4 (Pritchard et al. 2000). An admixture model with correlated allele frequencies without prior population information was used. A burn-in period of 250,000 was used, followed by 500,000 Markov Chain Monte Carlo (MCMC) permutations. Ten replicates (runs) were carried out for each possible value of K (from 1 to 10). Two different approaches were used to detect the most likely K value: the first was that proposed by Pritchard et al. (2000) and the second proposed by Evanno et al. (2005), using the web based Structure Harvester v.0.6.92 (Earl and VonHoldt 2012). Hierarchical distribution of genetic variation within and among groups of annatto accessions was evaluated using "locus-by-locus" AMOVA with GenAlEx version 6.5 (Peakall and Smouse 2012), with individuals clustered into groups according to the Structure analysis. Significance was assessed by conducting 10,000 permutations.

Results

Phytochemical compound characterization

Seed moisture content of samples was used to achieve uniformity of other parameters for correlation between different samples. Moisture content ranged from 5.78 to 14.32 g (100 g DM)$^{-1}$ (Table 1). Lipid concentrations ranged from 2.14 g to 7.11 g (100 g DM)$^{-1}$. Total carotenoids, expressed as bixin, showed concentrations (dry basis) ranging from a minimum of 2.00 g to a maximum of 7.31 g (100 g DM)$^{-1}$.

Analyses of tocotrienols identified the predominance of γ-tocotrienol and δ-tocotrienol, with δ-tocotrienol representing approximately 90% of the observed isoforms. The α and β-tocotrienols were not present or were below the analytical sensitivity limit of the method used [0.01 g (100 g DM)$^{-1}$]. Total tocotrienols concentrations (dry basis) ranged from a minimum of 0.25 g to a maximum of 1.05 g (100 g DM)$^{-1}$ (Table 1). Analyses of geranylgeraniol showed results ranging from 0.49 g to 2.61 g (100 g DM)$^{-1}$.

The cluster analysis conducted with phytochemical compounds classified the accessions into seven groups (Fig. 2). The accessions from São Paulo, Rio de Janeiro and Mato Grosso all clustered in the first three groups, which together showed low values for all traits. The accessions from Rondônia clustered in the other four groups, with the exception of accessions 63 (group 1) and 54 (group 2). These groups also contained one accession from Minas Gerais and eight of unknown origin. These four groups showed higher values than the other three groups for all traits. Moreover, within these four groups there are accessions individually responsible for the highest values in all characteristics, such as accessions 62 (Bixin: 7.11), 56 (Geranylgeraniol: 2.61), 47 and 23 (Tocotrienol: 1.05).

The PCA explained nearly 83% of the variation in the data set (Fig. 3). PCA1 can be considered an axis of phytochemical constituent concentration, with higher values to the left, attributed mostly to Rondônia accessions, while PCA2 distinguishes between annattos with more geranylgeraniol and lipids above the axis, and those with more tocotrienols and bixin below the axis.

Significant linear correlations were observed ($P < 0.05$) between the geranylgeraniol concentration and lipids ($r = 0.63$), bixin and lipids ($r = 0.51$), tocotrienols and lipids ($r = 0.41$) and between bixin and tocotrienols ($r = 0.56$). There were no correlations between geranylgeraniol and bixin, nor between geranylgeraniol and tocotrienols.

Molecular characterization

The characterization of 63 accessions with 16 microsatellite loci identified a total of 73 alleles, varying from two to ten alleles per locus, with an average of 4.56 alleles per locus (Table 2). We found

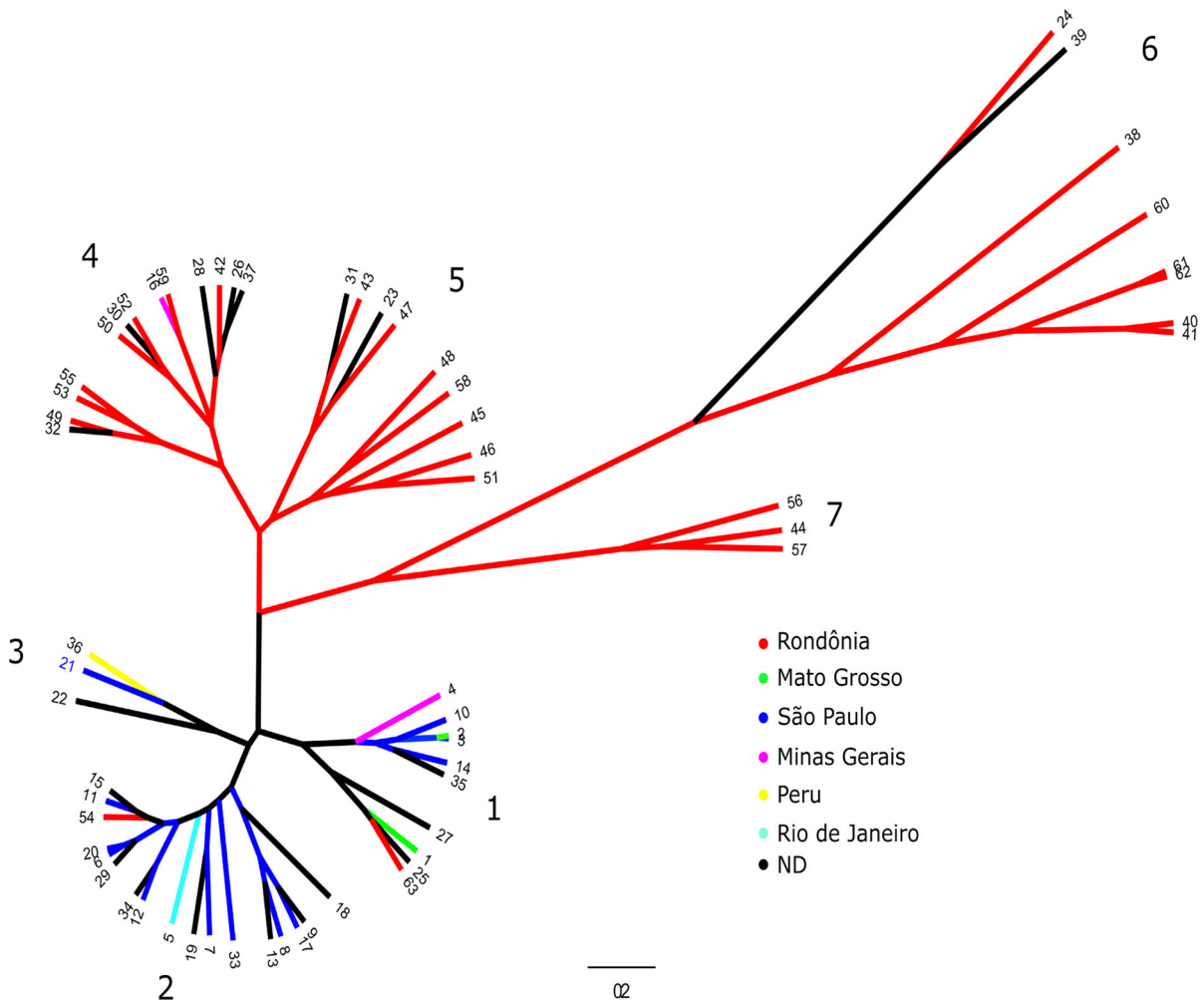

Fig. 2 Cluster analysis using Euclidean distances and the UPGMA algorithm based on four phytochemical compounds analyzed in 63 *Bixa orellana* accessions from the Brazilian States of Rondônia, Mato Grosso, São Paulo, Minas Gerais, Rio de Janeiro, from Peru and 20 with unknown origins (ND)

higher H_E then H_O values for all loci, indicating heterozygote deficits. As a result, F_{IS} was high for most loci and 14 loci were not in Hardy–Weinberg Equilibrium. This result is expected for genebanks where populations are not sampled, but individuals from diverse populations are grouped in a common area, disagreeing with the Hardy–Weinberg premises that population should be panmictic and with infinite size. Polymorphism Information Content (PIC) varied widely, with an average of 0.43.

In the PCoA analysis, the first two principal coordinates explained 48.3% of total variation and showed that accessions from Southeast and Central–West Brazil tend to form a separate group from the North (Rondônia) accessions, which were more dispersed and clustered mainly towards the right side of the first principal coordinate (Fig. 4). Cluster analysis based on Rogers' distance and the Neighbour-Joining algorithm (Fig. 5) presented two major clusters, revealing the same pattern observed in the PCoA. All accessions from Rondônia State, except two, were grouped close to each other, while accessions from Southeast Brazil formed a distinct group. The two Central–West (Mato Grosso) accessions were distributed in both major clusters. Bayesian analysis performed in Structure (Fig. 6) also confirmed the results obtained with PCoA and the NJ dendrogram. Two genetic clusters were obtained ($\Delta K = 776.35$), according to the Evanno method (S1 Fig.), showing that most of the accessions in the red group belong to

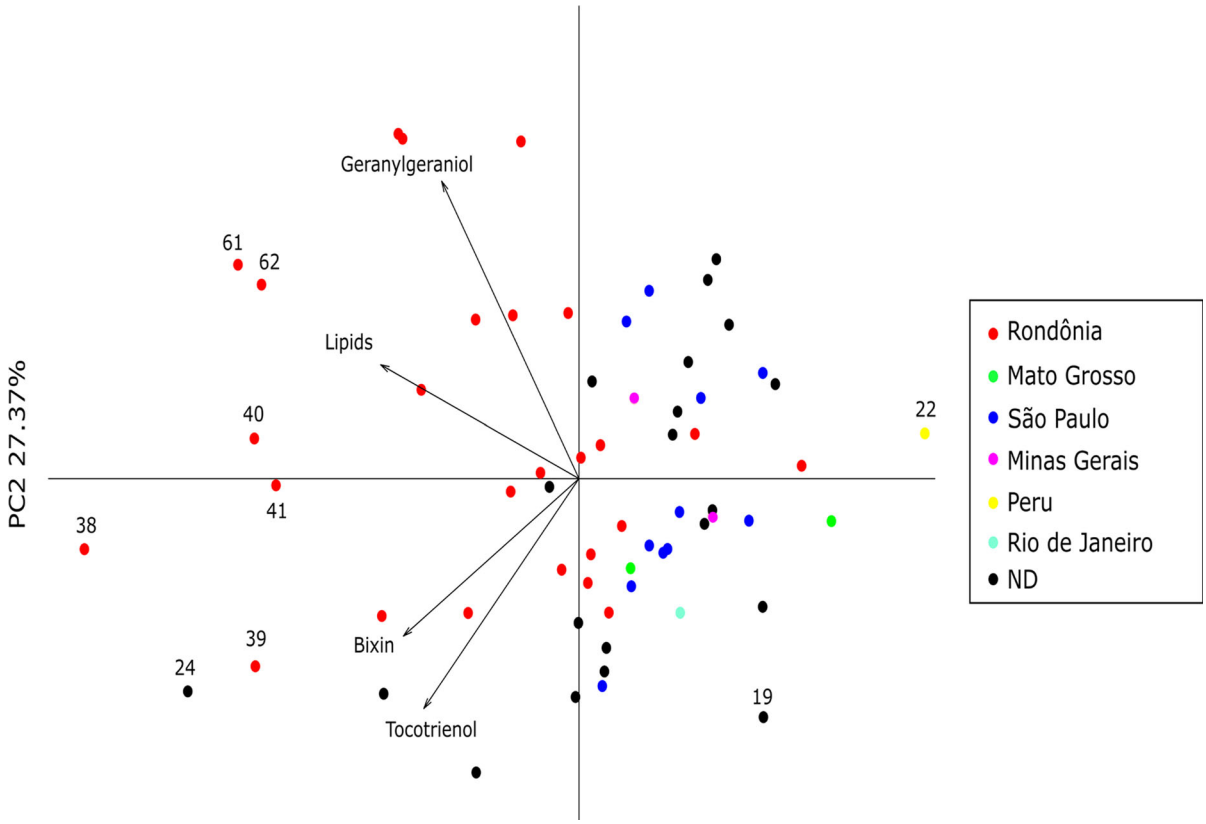

Fig. 3 Principal component analysis biplot of the 63 *Bixa orellana* accessions based on bixin, tocotrienol, geranylgeraniol and lipids analysis

the Southeast region, while those of the yellow group, with three exceptions, and excluding Peru and the non-determined origin accessions, were from northern Rondônia.

Results of AMOVA, using individuals clustered into groups according to the Structure analysis, showed that most of the genetic variation resided within groups (89%), while 11% of the variation resided between groups. Divergence was moderate between clusters ($F_{ST} = 0.112$, $P = 0.000$). The results of the PCoA, NJ tree and Structure showed signs of admixture (Figs. 4, 5, 6).

Discussion

Phytochemical characterization

The annatto accessions analyzed in this study showed moisture content values ranging from 5.78 to 14.32 g (100 g DM)$^{-1}$. Stringheta and Silva (2008) established 14 g (100 g)$^{-1}$ as the maximum moisture content of annatto seeds, above which there was a risk of mold growth. Only one accession (sample 45) showed a value above 14 g (100 g DM)$^{-1}$, which shows that the analyzed accessions present adequate moisture contents. As for the lipid contents, the analyzed accessions showed lipid values in agreement with Frega et al. (1998) and Rao et al. (2015), reporting an average of 5.3 (100 g DM)$^{-1}$ and 6.3 g (100 g DM)$^{-1}$, respectively.

Currently, the main criterion of seed quality used for the improvement of this crop is the concentration of bixin, which can be as high as 5.0 g (100 g DM)$^{-1}$ (Vilar et al. 2014). According to de Franco et al. (2008), levels higher than 2.5 g (100 g DM)$^{-1}$ are usually required for export. Annatto accessions maintained by this germplasm bank showed bixin concentrations ranging from 2.0 to 7.1 g (100 g DM)$^{-1}$, where most of the accessions (93.6%) produced values higher than 2.5 g (100 g DM)$^{-1}$, and therefore have export potential. This wide variation in bixin content

Table 2 Genetic parameters for the microsatellite analysis of *Bixa orellana* accessions from the Instituto Agronômico germplasm bank with 16 SSR loci: *A*, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; F_{IS}, inbreeding coefficient; PIC, polymorphism information content

SSR Loci	*A*	H_O	H_E	F_{IS}	PIC
BorA2	3	0.16	0.61	0.73	0.53
BorA3	3	0.12	0.35	0.65	0.30
BorA5	3	0.08	0.18	0.53	0.16
BorB12	4	0.22	0.52	0.56	0.46
BorF9	8	0.28	0.76	0.62	0.72
BorG11	7	0.29	0.76	0.62	0.72
BorB1	6	0.24	0.63	0.62	0.58
BorB4	3	0.17	0.51	0.66	0.39
BorB10	10	0.31	0.83	0.63	0.80
BorC12	3	0.10	0.37	0.74	0.31
BorD1	3	0.27	0.37	0.28	0.31
BorD2	2	0.00	0.32	1.00	0.26
BorG4	5	0.16	0.32	0.50	0.29
BorH3	2	0.03	0.03	0.01	0.03
BorH7	5	0.13	0.45	0.71	0.43
BorH10	6	0.24	0.57	0.58	0.53
Overall mean	4.56	0.17	0.47	0.63	0.43

in different varieties of annatto was also observed by Matos et al. (1992), Carvalho et al. (2010) and Mantovani et al. (2013).

Total tocotrienols concentration of these annatto accessions, ranging from 0.25 to 1.05 g (100 g DM)$^{-1}$, were much higher than the mean concentration of 0.14 g (100 g DM)$^{-1}$ reported by Frega et al. (1998). This variability shows the importance and viability of conducting studies to select varieties with high concentrations of tocotrienols. Geranylgeraniol is the major oily constituent of annatto seeds, representing 1% of dry seeds (Vilar et al. 2014). Analyses of geranylgeraniol in annatto seeds in this study showed values ranging from 0.49 to 2.61 g (100 g DM)$^{-1}$. Most of the accessions showed values well above those obtained by Smith and Wallin (2006) [average of 0.75 g (100 g DM)$^{-1}$]. As in the case of tocotrienols, the wide variation of the concentration of geranylgeraniol present in the annatto accessions in this germplasm collection indicates the need to conduct studies selecting varieties with a high concentration of this component.

This study showed that the Northern accessions generally have high concentrations of the phytochemical compounds, which leads us to speculate that this

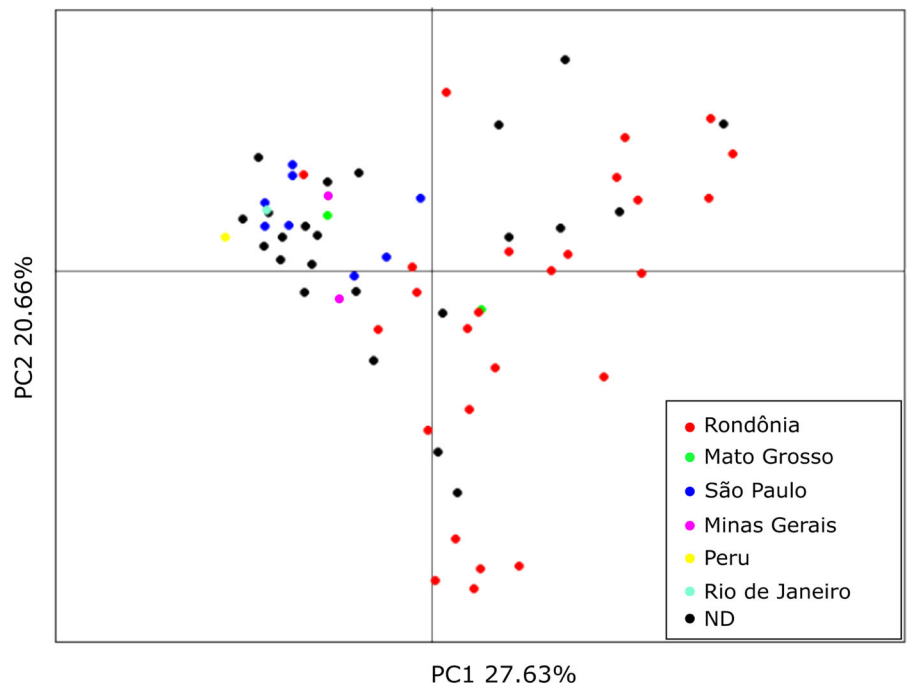

Fig. 4 Principal coordinate analysis of the dispersion of the 63 *Bixa orellana* accessions evaluated with 16 SSR markers

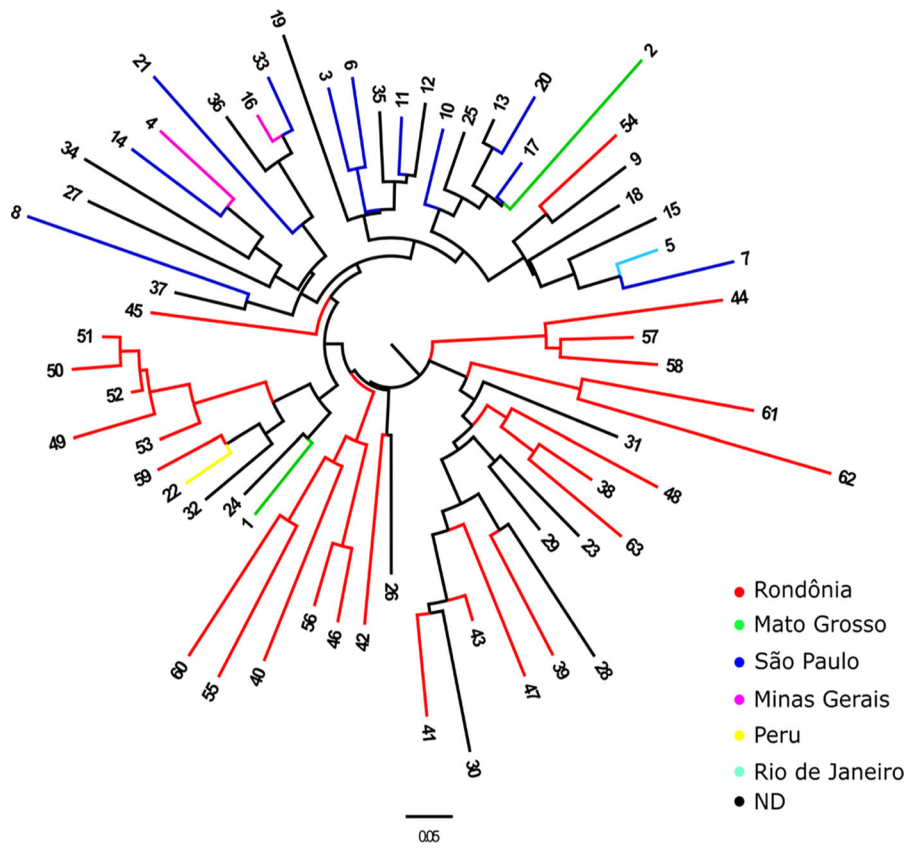

Fig. 5 Neighbour-Joining tree of 63 *Bixa orellana* accessions (*colored by origin*), constructed with Rogers' distance based on 16 SSR loci

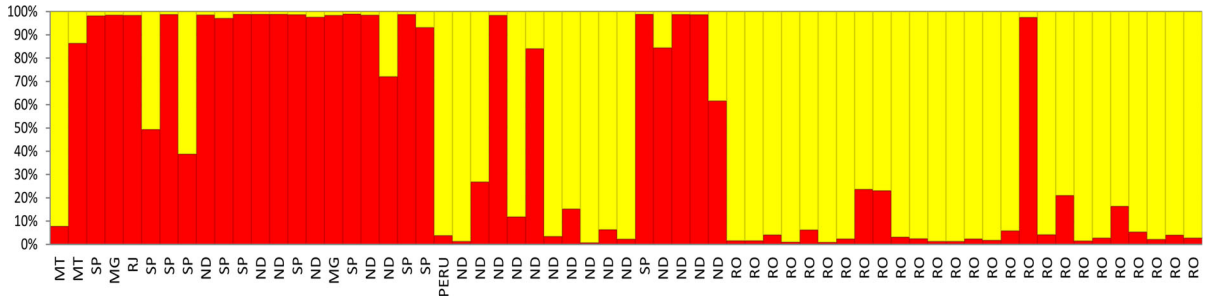

Fig. 6 The estimated proportion of membership in the corresponding clusters (K = 2), calculated using Structure, for 63 accessions of *Bixa orellana* evaluated with 16 SSR loci

finding could be related to the Amazonian origin and domestication of annatto (Sandy-Cuen and Becerra 2003; Clement et al. 2010; Moreira et al. 2015), yet to be confirmed. Accession 63, which morphologically appears to be a wild accession, presenting small fruits, fewer seeds per fruit and seeds only partially covered with aril, presented low values for all phytochemical characteristics, consistent with what would be expected for wild accessions. Due to the fact that wild accessions have not suffered human selection for traits of human interest, it is expected that it should present lower values compared to the domesticated materials (Moreira et al. 2015).

Molecular characterization

Heterozygosity is a measure of the genetic variability that estimates how much variation exists in a

population and how this variation is distributed depending on the alleles present at a given locus (Frankham et al. 2004). The expected heterozygosity values, with a mean of 0.47, exceeded those obtained for the observed heterozygosity, averaging 0.17, for all loci in this study. When plants from different populations are grouped and analyzed in a single group, as with accessions from germplasm banks, the expected heterozygosity is usually greater than the observed heterozygosity, which is known as the Wahlund effect (Hartl and Clark 1998). High values of F_{IS} observed in all loci analyzed allowed the interpretation that the genotypes studied are also subjected to inbreeding. From these values, we calculated the apparent outcrossing rate ($\hat{t}_a$), yielding a value of 0.22. From this value, it is possible to interpret that the apparent selfing rate of the plants collected for representation in the germplasm bank is 0.78. The existence of crosses between closely related individuals generates inbreeding and, therefore, increases homozygosity (Ritland 2002). These results are in agreement with those obtained by Rivera-Madrid et al. (2006), conducting controlled pollinations in annatto accessions in an experimental field in Mexico, suggesting that annatto can tolerate both types of pollination, with recorded cross-pollination rates of 57% and self-fertilization of 31.4%. Vilares et al. (1992) also concluded that there was natural selfing in annatto. It is worth mentioning that apparent outcrossing rate and apparent selfing rate are estimates, and do not replace progeny tests to determine the mating system, and therefore need further studies to confirm these data.

In all the molecular analyses, such as the cluster and PCoA analyses, the Rondônia accessions, with only a few exceptions, were classified into groups separated from the Southeastern accessions of Brazil, similar to the phytochemical characterization. They were also clearly classified in one group separated from the other accessions from Brazil in the Bayesian analysis, which classified the accessions into two groups, showing genetic differentiation among the accessions in Brazil, with a genetic structuring separating the Northern from the Southeastern accessions. These results are in agreement with those obtained by Carvalho et al. (2005), which presented a clear differentiation between Northern accessions from those obtained in other regions of the country. Annatto is native to the Americas, and most probably the Amazon region (Sandy-Cuen and Becerra 2003; Clement et al. 2010), which might explain the higher levels observed for the phytochemical traits and genetic diversity values from the Northern (Rondônia) accessions.

In conclusion, our data revealed high variation of phytochemical compounds in the 63 accessions from the annatto Germplasm Bank of IAC, corroborated by high levels of genetic diversity revealed by 16 SSR markers. Interestingly, Northern accessions concentrate higher concentrations of the phytochemical compounds and higher levels of genetic diversity. Some accessions presented bixin levels well above average. Considering that this trait is a key feature in this crop, these materials are very promising to be used in genetic improvement programs. Additional studies with local varieties, as well as wild populations of annatto collected in several regions in Brazil, are underway by our team and they should contribute to a better understanding of the distribution of genetic diversity in this country, and also further information concerning the center of domestication of *B. orellana*.

Acknowledgements This project was financially supported by São Paulo Research Foundation—FAPESP (2011/10357-8; 2012/08307-5). G. D. was supported by FAPESP with MSc and PhD scholarships (2011/05243-3; 2013/08884-5). E.A.V. and C.R.C. were supported by National Council for Scientific and Technological Development (CNPq) research fellowships. Authors wish to thank Instituto Agronômico—IAC/APTA, for providing the annatto samples from the annatto Germplasm Bank for this study, and Polo Regional Centro Norte—PRDTA/APTA in Pindorama-SP, for maintaining the annatto Germplasm Bank.

Compliance with ethical standards

Conflicts of interest The authors declare that there are no conflicts of interest.

References

Akshatha V, Giridhar P, Ravishankar GA (2011) Morphological diversity in *Bixa orellana* L. and variations in annatto pigment yield. J Hortic Sci Biotechnol 86:319–324. doi:10.1080/14620316.2011.11512767

Albuquerque CLC, Meireles MAA (2012) Defatting of annatto seeds using supercritical carbon dioxide as a pretreatment for the production of bixin: experimental, modeling and economic evaluation of the process. J Supercrit Fluids 66:86–95. doi:10.1016/j.supflu.2012.01.004

Alves RW, de Souza AAU, Souza SMAGU, Jauregi P (2006) Recovery of norbixin from a raw extraction solution of annatto pigments using colloidal gas aphrons (CGAs). Sep

Purif Technol 48:208–213. doi:10.1016/j.seppur.2005.07.014

Arce J (1999) El achiote *Bixa orellana* L: cultivo promisorio para el trópico. Earth, Turrialba

Auttachoat W, Germolec DR, Smith MJ et al (2011) Contact sensitizing potential of annatto extract and its two primary color components, cis-bixin and norbixin, in female BALB/c mice. Food Chem Toxicol 49:2638–2644. doi:10.1016/j.fct.2011.07.009

Baer DF (1976) Systematics of the genus *Bixa* and geography of the cultivated annatto tree. PhD thesis, University of California

Carvalho PRN, Carvalho CRL, Mantovani DMB (1991) Estudo da composição de sementes, cachopas, folhas e galhos do urucueiro. In: Proceedings of the seminário internacional de Corantes naturais para alimento. p 317

Carvalho PRN, da Silva MG, Moreira CGC (1993) Avaliação dos métodos espectrofotométricos de análise de sementes de urucum (*Bixa orellana* L.). Colet Inst Tecnol Aliment 23:181–188

Carvalho PRN, da Silva MG, Fabri EG et al (2010) Concentração de bixina e lipidios em sementes de urucum da coleção do Instituto Agronômico (IAC). Bragantia 69:519–524

Clement CR, de Cristo-Araújo M, d'Eeckenbrugge GC et al (2010) Origin and domestication of native Amazonian crops. Diversity 2:72–106. doi:10.3390/d2010072

Craveiro AR, Oliveira CLA, Araujo FWL (1989) The presence of geranylgeraniol in *Bixa orellana* Linn. Quim Nova 12:297–298

de Carvalho JFRP, Robinson IP, Alfenas AC (2005) Isozymic variability in a Brazilian collection of annatto (*Bixa orellana* L.). Pesqui Agropecu Bras 40:653–660. doi:10.1590/S0100-204X2005000700005

Dequigiovanni G, Ramos SLF, Zucchi MI et al (2014) Isolation and characterization of microsatellite loci for *Bixa orellana*, an important source of natural dyes. Genet Mol Res 13:9097–9102. doi:10.4238/2014.October.31.25

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dray S, Dufour A-B (2007) The ade4 Package: implementing the duality diagram for ecologists. J Stat Softw 22:1–20. doi:10.18637/jss.v022.i04

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. doi:10.1007/s12686-011-9548-7

Evanno 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. doi:10.1111/j.1365-294X.2005.02553.x

Fabri EG (2015) Demanda por corantes naturais aquece mercado brasileiro de urucum. Sociedade Nacional de Agricultura. http://sna.agr.br/demanda-por-corantes-naturais-aquece-mercado-brasileiro-de-urucum. Accessed 26 Nov 2015

Forzza RC et al (2010) Lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro, Rio de Janeiro

Franco CFO, Fabri EG, Barreiro Neto M et al (2008) Urucum: sistemas de produção para o Brasil. EMEPA, João Pessoa

Frankham R, Ballou JD, Briscoe DA (2004) A primer of conservation genetics. Cambridge University Press, Cambridge

Frega N, Mozzon M, Bocci F (1998) Identification and estimation of tocotrienols in the annatto lipid fraction by gas chromatography-mass spectrometry. J Am Oil Chem Soc 75:1723–1727. doi:10.1007/s11746-998-0323-1

Giorgi A, De Marinis P, Granelli G et al (2013) Secondary metabolite profile, antioxidant capacity, and mosquito repellent activity of *Bixa orellana* from Brazilian Amazon region. J Chem. doi:10.1155/2013/409826

Giuliano G, Rosati C, Bramley PM (2003) To dye or not to dye: biochemistry of annatto unveiled. Trends Biotechnol 21:513–516. doi:10.1016/j.tibtech.2003.10.001

Hartl D, Clark A (1998) Principles of population genetics. Sinauer Associates, Sunderland

Horwitz W (2005) Official methods of analysis of the association of official analytical chemists, 18th edn. AOAC, Gaithersburg

Jondiko IJO, Pattenden G (1989) Terpenoids and an apocarotenoid from seeds of *Bixa orellana*. Phytochemistry 28:3159–3162. doi:10.1016/0031-9422(89)80298-5

Kalia RK, Rai MK, Kalia S et al (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334. doi:10.1007/s10681-010-0286-9

Keenan K, McGinnity P, Cross TF et al (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol Evol 4:782–788. doi:10.1111/2041-210X.12067

Langella O (2002) Populations, 1.2.32. Population genetic software. http://bioinformatics.org/~tryphon/populations. Accessed 30 Aug 2016

Leal F, Clavijo CM (2010) Acerca de la história, taxonomia, botánica y usos de *Bixa orellana* L. Rev Unell Cienc Tec 1:78–86

Leal F, de Clavijo CM (2012) Annatto: botany and horticulture. In: Janick J (ed) Horticultural Reviews. Wiley, Hoboken, p 389–419

Lopes MV, Desoti VC, Caleare ADO et al (2012) Mitochondria superoxide anion production contributes to geranylgeraniol-induced death in leishmania amazonensis. Evid Based Complement Alternat Med 2012:298320. doi:10.1155/2012/298320

Mantovani NC, Grando MF, Xavier A, Otoni WC (2013) Avaliação de genótipos de urucum (*Bixa orellana* L.) por meio da caracterização morfológica de frutos, produtividade de sementes e teor de bixina. Cienc Florest 23:344–362. doi:10.5902/198050989281

Matos FJA, Alencar JW, Craveiro AA, Machado MIL (1992) Acidos graxos de algumas oleaginosas tropicais em ocorrência no nordeste brasileiro. Quim Nova 15:181–195

Monzote L, Montalvo AM, Almanonni S et al (2006) Activity of the essential oil from *Chenopodium ambrosioides* grown in Cuba against *Leishmania amazonensis*. Chemotherapy 52:130–136. doi:10.1159/000092858

Moreira PA, Lins J, Dequigiovanni G et al (2015) The domestication of annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. Econ Bot 69:127–135. doi:10.1007/s12231-015-9304-0

Nisar N, Li L, Lu S et al (2015) Carotenoid metabolism in plants. Mol Plant 8:68–82. doi:10.1016/j.molp.2014.12.007

Panfili G, Fratianni A, Irano M (2003) Normal phase high-performance liquid chromatography method for the determination of tocopherols and tocotrienols in cereals. J Agric Food Chem 51:3940–3944. doi:10.1021/JF030009V

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

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

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/

Rambaut A, Drummond A (2010) FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, UK. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 2 Apr 2013

Rao PP, Rao NG, Jyothirmayi T et al (2015) Characterisation of seed lipids from *Bixa orellana* and *Trachyspermum copticum*. J Am Oil Chem Soc 92:1483–1490. doi:10.1007/s11746-015-2717-1

Ritland K (2002) Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221–228. doi:10.1038/sj.hdy.6800029

Rivera-Madrid R, Escobedo-GM RM, Balam-Galera E et al (2006) Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). Sci Hortic 109:165–172. doi:10.1016/j.scienta.2006.03.011

Rodrigues J (1995) Análise de isoenzimas em progênies de meio-irmãos de urucum (*Bixa orellana* L.). Federal University of Viçosa, Viçosa

Rogers J (1972) Measures of genetic similarity and genetic distances. In: Studies in genetics VII. University of Texas Publication, Austin, pp. 145–153

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sandy-Cuen PM, Becerra R (2003) Manejo campesino de recursos naturales. El achiote. BioDiversitas 7:7–11

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234. doi:10.1038/72708

Sen CK, Khanna S, Roy S (2007) Tocotrienols in health and disease: the other half of the natural vitamin E family. Mol Aspects Med 28:692–728

Smith J, Wallin H (2006) Annatto extracts. Chemical and technical assessment ftp://ftp.fao.org/ag/agn/jecfa/cta_annatto.pdf.

Stringheta CP, Silva PI (2008) Pigmentos de urucum. Extração, reações químicas, uso e aplicações, 1st edn. Suprema, Viçosa

Tan B, Foley J (2000) Tocotrienols and geranylgeraniol from *Bixa orellana* byproducts. WO Pat. US6350453 B1

Vieira MLC, Santini L, Diniz AL, Munhoz CF (2016) Microsatellite markers: what they mean and why they are so useful. Genet Mol Biol 39:312–328. doi:10.1590/1678-4685-GMB-2016-0027

Vilar DA, Vilar MSA, de Lima e Moura TFA et al (2014) Traditional uses, chemical constituents, and biological activities of *Bixa orellana* L.: a review. Sci World J 857292:1–11. doi:10.1155/2014/857292

Vilares AS, São José AR, Rebouças TNH, Souza IVB (1992) Estudo da biologia floral de urucuzeiro (*Bixa orellana* L.). Rev Bras Corantes Nat 1:101–105

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, Sunderland

Zahn TJ, Eilers M, Guo Z et al (2000) Evaluation of isoprenoid conformation in solution and in the active site of protein-farnesyl transferase using carbon-13 labeling in conjunction with solution- and solid-state NMR. J Am Chem Soc 122:7153. doi:10.1021/ja000860f