New microsatellite loci for annatto (*Bixa orellana*), a source of natural dyes from Brazilian Amazonia

Gabriel Dequigiovanni[1*], Santiago Linorio Ferreyra Ramos[2], Maria Teresa Gomes Lopes[3], Charles Roland Clement[4], Doriane Picanço Rodrigues[3], Eliane Gomes Fabri[5], Maria Imaculada Zucchi[6] and Elizabeth Ann Veasey[1]

Abstract: *Annatto (Bixa orellana) is a tropical crop native to the Americas with Amazonia as the likely center of origin of domestication. Annatto is important because it produces the dye bixin, which is widely used in the pharmaceutical, food, cosmetic and textile industries. A total of 32 microsatellite loci were isolated from a microsatellite-enriched genomic library, of which 12 polymorphic loci were used to characterize four populations of B. orellana and B. orellana var. urucurana, the wild relative. Higher genetic diversity estimates were detected for the wild populations when compared to the cultivated populations. Also, higher apparent outcrossing rates were found for the two wild than the cultivated populations. These results indicate a mixed mating system for the species. All markers described herein have potential to be used in further studies evaluating the genetic diversity, population dynamics, domestication, breeding, and conservation genetics of annatto.*

Key words: *Bixaceae, conservation genetics, genetic diversity, population genetics, SSR.*

INTRODUCTION

Annatto (*Bixa orellana* L.; Bixaceae) is a tropical crop native to the Americas with Amazonia as the likely center of origin of domestication (Arce 1999, Clement et al. 2010, Moreira et al. 2015)domestication and dispersal of native Amazonian crops in an expanding archaeological context. Solid molecular data are available for manioc (Manihot esculenta. The wild ancestor of cultivated annatto has recently been identified as *B. orellana* var. *orellana* (Willd.) Kuntze ex Pilg. Hence (Moreira et al. 2015). Historical evidence indicates extensive distribution and cultivation of annatto in the American tropics and subsequently its spread worldwide (Leal and Clavijo 2010). Brazil is the major producer of annatto and also hosts the greatest diversity of this species. Annatto is also produced in Peru, Kenya, the Dominican Republic, Colombia, Jamaica, Costa Rica, Suriname, and other countries in Asia (Akshatha et al. 2011). Annatto is commercially valuable due to its applications in the food and cosmetics industries, as a natural substitute of synthetic dyes (Nisar et al. 2015). It is the second most important economic crop worldwide among all natural colorants, and has gained fame for containing, apart from dye, other important substances for human

*Corresponding author:
E-mail: gabriel.dequi@gmail.com

health, e.g., geranylgeraniol, tocotrienols and other carotenoids with antimicrobial, antioxidant and antiviral properties (Albuquerque and Meireles 2012)supercritical CO 2 extraction for defatting of annatto seeds was studied; the objective was to obtain an extract rich in tocotrienols and the defatted rich-bixin seeds. The process conditions were selected from global yield isotherms assays performed at 313 and 333 K, and 20, 31, and 40 MPa; the ratio of solvent mass (S, which can be used to treat human diseases, including leishmaniosis (Lopes et al. 2012). Annatto dye also plays a role in the Brazilian culture, since it is still used by indigenous tribes for body painting and dyeing of clothes (Plotkin 1993).

Microsatellites or simple sequence repeats (SSR) are important tools to assess the genetic diversity and genetic structure of populations. They are widely present in eukaryotic genomes and very useful, mainly because of their codominant inheritance, high polymorphism, high variability and suitability for automated allele sizing and cross-species transferability (Kalia et al. 2011, Vieira et al. 2016)using both low and high throughput genotyping approaches. Motivated by the importance of these sequences over the last decades this review aims to address some theoretical aspects of SSRs, including definition, characterization and biological function. The methodologies for the development of SSR loci, genotyping and their applications as molecular markers are also reviewed. Finally, two data surveys are presented. The first was conducted using the main database of Web of Science, prospecting for articles published over the period from 2010 to 2015, resulting in approximately 930 records. The second survey was focused on papers that aimed at SSR marker development, published in the American Journal of Botany's Primer Notes and Protocols in Plant Sciences (over 2013 up to 2015. Dequigiovanni et al. (2014)while 15 were considered monomorphic. The mean number of alleles per locus was 3.8, ranging from 2 to 6 alleles per locus. Mean values for the observed and expected heterozygosities were 0.541 (ranging from 0 to 0.658 developed 10 polymorphic microsatellite markers for *B. orellana* L.; however, a larger number of markers can increase the accuracy of estimates of   population genetic parameters. Thus, this study presents a new set of microsatellite loci for *B. orellana* and its wild relative *B. orellana* var. *urucurana*, with a view to generate useful information for conservation strategies and population genetic studies.

MATERIAL AND METHODS

Ninety-eight individuals from four populations were analyzed in this study. Two were populations of cultivated annatto (*B. orellana*), one with 31 accessions located in São Francisco do Guaporé (lat 12° 16' 08.2" S, long 63° 25' 27.1" W, alt 176 m asl), Rondônia, Brazil, and the other with 22 accessions located in Rondon do Pará (lat 4° 44' 43.4" S, long 47° 55' 59.5" W, alt 165 m asl), Pará, Brazil. Markers developed in this study were also tested for cross amplification in two wild annatto (*B. orellana* var. *urucurana*) populations, one from Corumbiara (lat 13° 39' 15.80" S, long 61° 32' 27.69" W, alt 340 m asl), Rondônia, Brazil, with 25 individuals, and the other from Ariquemes (lat 9° 55' 30.5" S, long 63° 04' 16.6" W, alt 142 m asl), Rondônia, Brazil, with 20 individuals.

Genomic DNA extraction from *Bixa orellana* and *B. orellana* var. *urucurana* samples was performed with the CTAB protocol (Doyle and Doyle 1990). A microsatellite-enriched library for *B. orellana* was developed following Billotte et al. (1999). Genomic DNA was digested with the enzyme *Afa*I (Integrated DNA Technology-IDT, Coralville, USA) and the fragments resulting from digestion were linked to *Afa*21 and *Afa*25 adapters. Fragments were pre-amplified by Polymerase Chain Reaction (PCR) using the *Afa*21 adapter. Fragment-containing repeats were selected with $(CTT)_{10}$, $(GT)_{10}$ and $(TA)_{10}$ biotinylated oligos, and recovered with streptavidin-coated magnetic particles (Sigma-Aldrich, St. Louis, USA). Enriched DNA fragments were amplified and cloned using the pGEM-T easy vector (Promega, Madison, USA) and transformed into XL1-BLUE *Escherichia coli* competent cells (Stratagene, Santa Clara, USA). Ninety-two positive clones were sequenced using universal T7 and SP6 primers with a BigDye v3.1 terminator kit on an ABI 3130XL Genetic Analyzer automated sequencer (Applied Biosystems, Foster City, USA). The sequences containing microsatellite repeats were selected using WebSat (Martins et al. 2009). We considered dinucleotides with more than six repeats, and trinucleotides, tetranucleotides and pentanucleotides with three or more repeats. The primers were designed considering sequences with 50-80% of GC content in PRIMER 3 (Rozen and Skaletsky 2000), with final products ranging from 130 to 350 base pairs (bp) and primer size ranges from 18 to 22 bp. An M13 sequence tail was added to the 5' end of each forward primer following the Schuelke (2000) protocol.

Thirty-two microsatellite loci were characterized. Polymerase chain reactions were performed in a final volume of 10 µL, containing 20 ng of genomic DNA template, 1 U *Taq* DNA polymerase (Fermentas, Vilnius, Lithuania), 1X PCR buffer (10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.08% Nonidet P40), 0.25 mM of 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. The PCRs were carried out according to the Schuelke (2000) protocol, consisting of 94 °C (5 min), then 30 cycles at 94 °C (30 s)/Ta °C (45 s)/72 °C (45 s) [Ta= annealing temperature (Table 1)], 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. The quality of amplification was checked by electrophoresis in agarose gels (1.5 %) stained with GelRed (Biotium, Hayward, USA). The PCR products were visualized in an ABI 3130*XL* sequencer (Applied Biosystems, Foster City, USA) and allele sizes were scored using *GeneScan™–500 ROX® Size Standard* (Applied Biosystems, Foster City, USA) and analyzed with GENEMAPPER v4.0 software (Applied Biosystems, Foster City, USA). Descriptive statistics and Hardy–Weinberg Equilibrium (HWE) were calculated using *diveRsity* (Keenan et al. 2013) for R (R Core Team 2015). Genotypic disequilibrium between pairwise loci was estimated using FSTAT (Goudet 2002). Monte Carlo permutations of alleles between individuals and a Bonferroni correction (95%; α=0.05) were used to test if the estimates were significantly different from zero. The software micro-checker 2.2.1 (van Oosterhout et al. 2004) was used to identify possible genotyping errors resulting from stuttering or large allele dropout and the presence of null alleles within the microsatellite data set by performing 1000 randomizations.

The distribution of genetic variation within and among populations was evaluated using "locus-by-locus" AMOVA with GenAlEx version 6.5 (Peakall and Smouse 2012). Wright's F_{ST} was also used to estimate population differentiation and was calculated using GenAlEx. When populations are under Wright's equilibrium, the outcrossing rate is a function of the within-population inbreeding coefficient (Wright 1965). So, the apparent outcrossing rate ($\hat{t}_a$) was calculated for all populations according to Vencovsky (1994), with $\hat{t}_a = (1 - f) / (1 + f)$. Principal coordinate analysis (PCoA) was used to evaluate the dispersion of accessions with GenAlEx (Peakall and Smouse 2012)enabling population genetic analyses of codominant, haploid and binary data. Allele frequency-based analyses include heterozygosity, F statistics, Nei\u2019s genetic distance, population assignment, probabilities of identity and pairwise relatedness. Distance-based calculations include AMOVA , principal coordinates analysis (PCA).

RESULTS AND DISCUSSION

Thirty-two loci were amplified successfully (Table 1) from 92 positive clones sequenced from the library. Of these 32 loci, 12 were found to be polymorphic in *B. orellana* and *B. orellana* var. *urucurana* populations (Table 2). This level of polymorphism (35%) was also observed in other studies with *Bixa orellana*. Ten polymorphic in 25 evaluated loci were detected by Dequigiovanni et al. (2014). Software Micro-Checker detected no genotyping errors due to stuttering and large allele dropout. The analyses also showed that loci BorA5_2013, BorB1_2013, BorD1_2013, BorD2_2013, BorG11_2013, and BorH10_2013 might be affected by null alleles in cultivated populations. This excess of homozygosity may be attributable to inbreeding. Therefore, none of the loci were excluded from the analyses.

Polymorphic loci were used to calculate descriptive statistics for each population (Table 2). For the wild *B. orellana* var. *urucurana,* the number of alleles per locus varied from 1 to 8. A lower number of alleles per loci was found for cultivated annatto, varying from 1 to 6 (Table 2). The average observed (H_O) and expected heterozygosities (H_E) were also higher in the wild than cultivated populations, with higher H_E than H_O values observed in both wild and cultivated populations. As a result, local inbreeding coefficients were high in all populations (Table 2). Similar results for cultivated accessions were reported by Dequigiovanni et al. (2014) (A=3.8; H_O=0.54; H_E=0.63).

Higher levels of genetic diversity in wild than cultivated crops were also found in other crops, due to genetic bottleneck effects during domestication, such as in tepary beans (*Phaseolus acutifolius) (Blair et al. 2012, Gujaria-Verma et al. 2016), common beans (P. vulgaris) (Bitocchi et al. 2013), apricot (Prunus armeniaca*) *(Bourguiba et al. 2012), and sunflower* (*Helianthus annuus*) *(Mandel et al. 2011). However, this is not always the case, since in some crops, e.g., in carrot (Daucus carota subsp. sativus), no decrease of genetic diversity occurred during domestication (Iorizzo et al. 2013).*

Deviation from Hardy–Weinberg equilibrium (HWE) was tested for all loci and populations. Ten loci were found deviating from HWE due to excess heterozygosity for *B. orellana* and five loci for *B. orellana* var. *urucurana*. Deviations from HWE may occur because *B. orellana* has a mixed mating system and can tolerate both autogamy and allogamy (Rivera-Madrid et al. 2006, Valdez-Ojeda et al. 2010, Joseph et al. 2012)changes in botanical composition could cause variable (unstable. Similarly, Dequigiovanni et al. (2014) found deviations from HWE in eight out of ten loci analyzed. No significant linkage disequilibrium was detected for any pair of loci tested after Bonferroni correction.

Table 1. Description of 32 *Bixa orellana* microsatellite loci, including loci names, GenBank accession numbers, annealing temperatures (T_a), repeat motifs and size range of each locus

Locus	GenBank accession	T_a (°C)	Repeat motif	Size range (bp)
BorA3_2013	KJ922640	60.0	$(TC)_{16}$	216-252
BorA5_2013	KJ922646	62.0	$(TG)_{13}(GA)_{22}$	237-263
BorB1_2013	KJ922641	62.0	$(TG)_8 N (TG)_{11}$	177-255
BorB5_2013	KJ922647	58.0	$(CA)_{19}$	259-291
BorC5_2013	KJ922648	60.0	$(CT)_{14}(CA)_9$	245-263
BorD1_2013	KJ922642	58.0	$(TG)_8(GA)_4$	167-177
BorD2_2013	KJ922643	62.0	$(TG)_{10}$	244-264
BorF9_2013	KJ922651	60.0	$(TA)_6(TG)_{10}$	290-310
BorG11_2013	KJ922655	56.0	$(TA)_6(CA)_{21}$	291-349
BorH3_2013	KJ922645	60.0	$(TG)_8$	236-244
BorH7_2013	KJ922652	56.0	$(AC)_8$	264-270
BorH10_2013	KJ922656	56.0	$(TA)_4(CA)_9$	295-327
BorC2_2013	KX017260	60.0	$(AG)_{30}$	209
BorG3_2013	KJ922644	58.0	$(GA)_{16}$	160
BorH1_2013	KX017261	62.0	$(AG)_{16} N (AG)_{10}$	205
BorC4_2013	KX017262	60.0	$(TTAT)_3$	151
BorE4_2013	KX017263	58.0	$(TC)_{10}(CA)_7$	194
BorF6_2013	KX017264	60.0	$(ATA)_4$	238
BorH6_2013	KJ922649	60.0	$(TC)_5(CA)_9$	206
BorB8_2013	KX017265	58.0	$(TG)_5$	204
BorC8_2013	KX017266	58.0	$(AG)_5$	202
BorD7_2013	KJ922650	58.0	$(GT)_9$	177
BorD9_2013	KX017267	58.0	$(AC)_8$	268
BorH9_2013	KX017268	60.0	$(AT)_5(GT)_7$	204
BorB11_2013	KX017269	60.0	$(AC)_{11}$	245
BorB12_2013	KX017270	60.0	$(AG)_{22}$	130
BorC10_2013	KJ922653	60.0	$(CA)_8$	222
BorC11_2013	KX017271	60.0	$(TG)_9(AG)_{17}$	272
BorE11_2013	KJ922654	60.0	$(CA)_8$	251
BorE12_2013	KX017272	60.0	$(AG)_{18}$	156
BorG10_2013	KX017273	58.0	$(GA)_{24}$	167
BorH11_2013	KX017274	58.0	$(CT)_{12}(AC)_9$	105

The apparent outcrossing rates estimated for all populations in this study indicated a mixed mating system for annatto, with much higher outcrossing rates observed for the two wild populations ($\hat{t}_a = 0.644$ for Corumbiara; $\hat{t}_a = 0.759$ for Ariquemes) than for cultivated annatto ($\hat{t}_a = 0.198$ for São Francisco do Guaporé/RO; $\hat{t}_a = 0.355$ for Rondón do Pará, PA). Also, it is interesting to mention that the commercial annatto plantation in São Francisco do Guaporé was much more uniform than the one in Rondón do Pará, on a more traditional type of farm, reflected in a lower outcrossing rate of the former.

The AMOVA analysis identified a higher proportion of genetic variation within (68%) than among populations (29%, $F_{ST}=0.317$, P<0.001), which is still quite high and suggested that subdivision has a great impact on the genetic diversity. However, only 2% of the total variation was attributable to differences between wild and cultivated populations, showing that there must be considerable gene flow between these two types of populations, especially in Rondônia (Figure 1). The *F*-statistics ($F_{IS}=0.366$; $F_{ST}=0.367$; $F_{IT}=0.597$) also confirmed high levels of genetic structure. The cultivated population of Rondon do Pará was the most divergent, apparently indicating isolation by distance, while considerable gene flow was detected in the two wild and one cultivated populations in Rondônia.

In conclusion, the 12 polymorphic loci reported in this study proved to be powerful tools for assessing genetic diversity, genetic structure, as well as for domestication studies of *B. orellana* and *B. orellana* var. *urucurana*. Higher

Table 2. Genetic characterization of 12 polymorphic SSR loci in *Bixa orellana* (cultivated) and *B. orellana* var. *urucurana* (wild) populations. Genetic diversity described as number of alleles (*A*), observed (H_O) and expected (H_E) heterozygosities and *inbreeding coefficient (f= 1 -H_O/H_E)*

	Wild - Corumbiara				Wild – Ariquemes			
Locus	*A*	H_O	H_E	*f*	A	H_O	H_E	f
BorA3_2013	6	0.636	0.742	0.142	5	0.600	0.729	0.177
BorA5_2013	3	0.300	0.515	0.417	6	0.722	0.725	0.004
BorB1_2013	5	0.333	0.597	0.442	8	0.450	0.733	0.386
BorB5_2013	3	0.304	0.328	0.072	3	0.200	0.184	-0.088
BorC5_2013	7	0.792	0.792	0.000	5	0.900	0.646	-0.393
BorD1_2013	1	-	-	-	2	0.400	0.320	-0.250
BorD2_2013	8	0.792	0.814	0.028	4	1.000	0.591	-0.691
BorF9_2013	6	0.263	0.781	0.663	4	0.143	0.311	0.541
BorG11_2013	8	0.458	0.827	0.446	8	0.400	0.780	0.487
BorH3_2013	2	0.167	0.153	-0.091	4	0.200	0.597	0.661
BorH7_2013	2	0.600	0.471	-0.273	3	0.450	0.626	0.281
BorH10_2013	7	0.292	0.628	0.536	2	0.230	0.500	0.540
Mean	4.84	0.448	0.604	0.216	4.50	0.474	0.561	0.137
Total	58	-	-	-	54	-	-	-
	Cultivated - São Francisco do Guaporé				Cultivated - Rondon do Pará			
Locus	*A*	H_O	H_E	*f*	A	H_O	H_E	f
BorA3_2013	5	0.455	0.684	0.335	3	0.278	0.329	0.155
BorA5_2013	3	0.194	0.629	0.692	2	0.143	0.278	0.486
BorB1_2013	6	0.226	0.713	0.683	3	0.400	0.629	0.364
BorB5_2013	4	0.133	0.336	0.603	4	0.316	0.582	0.457
BorC5_2013	5	0.308	0.553	0.444	2	0.381	0.444	0.143
BorD1_2013	1	-	-	-	2	0.000	0.165	1.000
BorD2_2013	2	0.000	0.391	1.000	1	-	-	-
BorF9_2013	4	0.182	0.498	0.635	5	0.438	0.678	0.354
BorG11_2013	5	0.120	0.730	0.836	2	0.000	0.100	1.000
BorH3_2013	1	-	-	-	1	-	-	-
BorH7_2013	3	0.133	0.472	0.717	1	-	-	-
BorH10_2013	5	0.133	0.526	0.746	3	0.381	0.571	0.333
Mean	3.67	0.188	0.553	0.669	2.41	0.259	0.419	0.476
Total	44	-	-	-	29	-	-	-
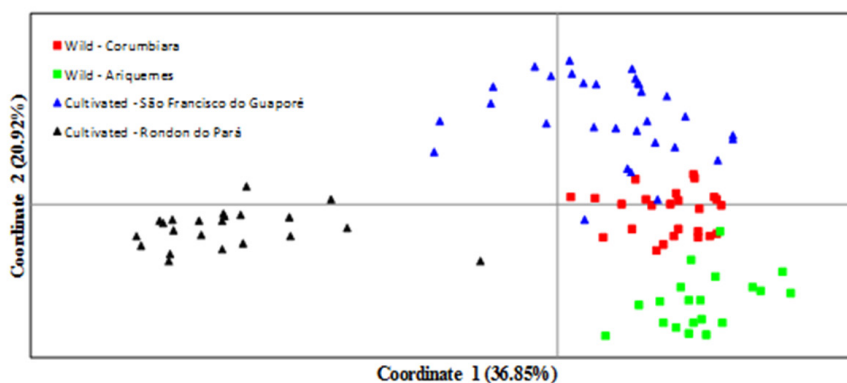
Figure 1. Principal coordinate analysis of the dispersion of two cultivated populations of *Bixa orellana* (Rondon do Pará and São Francisco do Guaporé) and two wild populations of *B. orellana* var. *urucurana* (Corumbiara and Ariquemes), using 12 newly developed microsatellite markers.

levels of genetic diversity and outcrossing rates were found for the wild than the cultivated populations. Also, most of the variation detected by SSR markers was located within populations, which apparently have a mixed mating system. Loci appearing as monomorphic in these populations may be classified as polymorphic in other populations and should therefore not be discarded.

ACKNOWLEDGEMENTS

REFERENCES

Akshatha V, Giridhar P and Ravishankar GA (2011) Morphological diversity in *Bixa orellana* L. and variations in annatto pigment yield. **Journal of Horticulture Science and Biotechnology 86**: 319-324.

Albuquerque CLC and 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. **Journal of Supercrit Fluids 66**: 86–94.

Arce J (1999) **El achiote *Bixa orellana* L. cultivo promisorio para el trópico.** Earth, Turrialba, 149p.

Billote N, Lagoda PJL, Risterucci AM and Baurens FC (1999) Microsatellite-enriched libraries: Applied methodology for the development of SSR markers in tropical crops. **Fruits 54**: 277-288.

Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti Zeuli P, Gioia T, Logozzo G, Attene G, Nanni L and Papa R (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. **New Phytologist 197**: 300-313.

Blair MW, Pantoja W and Carmenza Muñoz L (2012) First use of microsatellite markers in a large collection of cultivated and wild accessions of tepary bean (*Phaseolus acutifolius* A. Gray). **Theoretical and Applied Genetics 125**: 1137-1147.

Bourguiba H, Audergon JM, Krichen L, Trifi-Farah N, Mamouni A, Trabelsi S, D'Onofrio C, Asma BM, Santoni S and Khadari B (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. **BMC Plant Biology 12**: 49-64.

Clement CR, Cristo-Araújo M, d'Eeckenbrugge GC, Alves-Pereira A and Picanço-Rodrigues D (2010) Origin and domestication of native Amazonian crops. **Diversity 2**: 72-106.

Dequigiovanni G, Ramos SLF, Zucchi MI, Bajay MM, Pinheiro JB, Fabri EG, Bressan EA and Veasey EA (2014) Isolation and characterization of microsatellite loci for *Bixa orellana*, an important source of natural dyes. **Genetics and Molecular Research 13**: 9097-9102.

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. **Focus 12**: 13-15.

Goudet J (2002) FSTAT, a program to estmate and test gene diversites andfxaton indices (Version 2.9.3.2). Available at < htp://www2.unil.ch/popgen/sofwares/ fstat.htm>. Accessed in March, 2017.

Gujaria-Verma N, Ramsay L, Sharpe AG, Sanderson LA, Debouck DG, Tar 'an B and Bett KE (2016) Gene-based SNP discovery in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*) for diversity analysis and comparative mapping. **BMC Genomics 17**: 239-254.

Iorizzo M, Senalik DA, Ellison SL, Grzebelus D, Cavagnaro PF, Allender C, Brunet J, Spooner DM, Van Deynze A and Simon PW (2013) Genetic structure and domestication of carrot (*Daucus carota* subsp. *sativus*) (Apiaceae). **American Journal of Botany 100**: 930-938.

Kalia RK, Rai MK, Kalia S, Singh R and Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. **Euphytica 177**: 309-334.

Keenan K, McGinnity P, Cross TF, Crozier WW and Prodöhl PA (2013) diveRsity : An R package for the estimation and exploration of population genetics parameters and their associated errors. **Methods in Ecology and Evolution 4**: 782-788.

Leal F and Clavijo CM (2010) Acerca de la história, taxonomia, botánica y usos de *Bixa orellana* L. **Revista UNELLEZ de Ciencia y Tecnología 1**: 78-86.

Lopes MV, Desoti VC, Caleare ADO, Ueda-Nakamura T, Silva SO and Nakamura CV(2012) Mitochondria superoxide anion production contributes to geranylgeraniol-induced death in leishmania amazonensis. **Evidence-Based Complementary and Alternative Medicine 1**: 298320-298329.

Mandel JR, Dechaine JM, Marek LF and Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. **Theoretical and Applied Genetics 123**: 693-704.

Martins WS, Lucas DC, Neves KF and Bertioli DJ (2009) WebSat-a web software for microsatellite marker development. **Bioinformation 3**: 282-283.

Moreira PA, Lins J, Dequigiovanni G, Veasey EA and Clement CR (2015) The domestication of annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. **Economic Botany 69**: 127-135.

Nisar N, Li L, Lu S, Khin NC and Pogson BJ (2015) Carotenoid metabolism in plants. **Molecular Plant 8**: 68-82.

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

Plotkin M (1993) **Tales of a Shaman's apprentice**. Penguin Books, New York, 344p.

R Core Team (2015) **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna. Available at <https://www.r-project.org/>. Accessed in Oct, 2016.

Rivera-Madrid R, Escobedo RMM, Balam-Galera E, Vera-Ku M and Harries H (2006) Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). **Scientia Horticulturae 109**: 165-172.

Rozen S and Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. **Methods in Molecular Biology 132**: 365-386.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. **Nature Biotechnology 18**: 233-234.

Valdez-Ojeda R, Quiros CF, Aguilar-Espinosa ML and Rivera-Madrid R (2010) Outcrossing rates in annatto determined by sequence-related amplified polymorphism. **Agronomy Journal 102**: 1340-1345.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. **Molecular Ecology Notes 4**: 535-538.

Vencovski R (1994) Variance of an estimative of the outcrossing rate. **Revista Brasileira de Genética 17**: 349-351.

Vieira MLC, Santini L, Diniz AL and Munhoz CF (2016) Microsatellite markers: what they mean and why they are so useful. **Genetics and Molecular Biology 39**: 312-328.

Wright S (1965) The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. **Evolution 19**: 395.