GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Isolation and characterization of microsatellite loci for *Bixa orellana*, an important source of natural dyes

G. Dequigiovanni[1], S.L.F. Ramos[1], M.I. Zucchi[2], M.M. Bajay[1], J.B. Pinheiro[1], E.G. Fabri[3], E.A. Bressan[1] and E.A. Veasey[1]

[1]Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP, Brasil
[2]Agência Paulista de Tecnologia dos Agronegócios, Pólo Apta Centro Sul, Piracicaba, SP, Brasil
[3]Centro de Horticultura, Instituto Agronômico, Campinas, SP, Brasil

Corresponding authors: G. Dequigiovanni / E.A. Veasey
E-mail: gabriel.d@usp.br / eaveasey@usp.br

Genet. Mol. Res. 13 (4): 9097-9102 (2014)
Received March 24, 2014
Accepted August 8, 2014
Published October 31, 2014
DOI http://dx.doi.org/10.4238/2014.October.31.25

ABSTRACT. Annatto (*Bixa orellana*) is a plant native from the American continental tropical zone. The seeds are used to produce a carotenoid-based yellow to orange food coloring. Microsatellite markers were developed for the Brazilian native species *Bixa orellana* to describe its genetic diversity and structure as well as to support conservation studies. Twenty-five microsatellite loci were isolated and characterized using an enriched genomic library. Ten loci were polymorphic in the 50 accessions sampled in this study, 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) and 0.639 (ranging from 0.422 to 0.787), respectively. All markers described in this study will be useful in further studies evaluating the genetic diversity, population dynamics, and conservation genetics of

Bixa orellana.

Key words: *Annatto*; Bixaceae; Bixin; Genetic diversity; Simple-sequence repeat

INTRODUCTION

Annatto (*Bixa orellana* L.) is a tropical crop indigenous to America, with the Amazon region considered to be the center of diversity for the species (Arce, 1999; Clement et al., 2010). The major producers of annatto include Brazil, Peru, and Kenya, although annatto is also grown in the Dominican Republic, Colombia, Jamaica, Costa Rica, and Suriname and on a smaller scale in some countries in Asia (Akshatha et al., 2011).

The production of annatto in Brazil is around 12,000-13,000 t/year (IBGE, 2009), and the natural dye, known as bixin, is extracted from its seeds and used in the pharmaceutical, textile, dairy, food, beverage, paint, and cosmetics industries (Carvalho, 1992). Thus, as the only source of natural dye, this species is very important. *Bixa orellana* is also an important medicinal species (Lorenzi, 2002; Russell et al., 2005).

Little is known regarding the diversity and genetic structure of this species. Isozyme markers (Medina et al., 2001; Carvalho et al., 2005) have been used to study the genetic diversity among accessions from Venezuela and Brazil. DNA markers, such as sequence-related amplified polymorphism) (Valdez-Oheda et al., 2010) have also been used to examine the genetic diversity of annatto accessions from Mexico, and an amplified fragment length polymorphism protocol has also been established for this species (Fay et al., 2005). However, microsatellite markers have not been used to study the genetic diversity and structure of *B. orellana.*

Microsatellite markers or simple sequence repeats (SSR) feature high polymorphism, are co-dominant, and show highly reproducible results, making them a useful tool for assessing the genetic diversity and structure of plant populations (Kalia et al., 2012). In this study, we developed 25 SSR loci for *B. orellana* using an enriched DNA genomic library and examined the applicability of these primers in a group of annatto accessions from Brazil.

MATERIAL AND METHODS

Fifty accessions from the germplasm bank of *B. orellana* from the Instituto Agronômico (IAC) were sampled, originating from the states of Rondônia (26 accessions), São Paulo (5), Mato Grosso (2), and Minas Gerais (1) in Brazil, as well as 16 without a defined origin. A microsatellite-enriched genomic library was constructed using the procedure described by Billotte et al. (1999). Genomic DNA was extracted from recently expanded young leaves of *B. orellana* according to the method described by Doyle and Doyle (1990). To develop the genomic library, genomic DNA was extracted from 3 *B. orellana* accessions from the IAC germplasm bank.

Genomic DNA was digested with the enzyme *Afa*I. Fragments resulting from digestion were ligated to *Afa*21 and *Afa*25 adapters. Fragments were pre-amplified by polymerase chain reaction using the *Afa*21 adapter. The enrichment process was performed through the hybridization of biotinylated oligos [biotIII $(CTT)_{10}$, biotIII $(TA)_{10}$, and biotIII $(GT)_{10}$]. The fragments containing repeats were recovered and linked to streptavidin-coated magnetic particles. Enriched DNA fragments were amplified and cloned using the pGEM-T Easy vector

(Promega, Madison, WI, USA) and transformed into XL1-BLUE *Escherichia coli* competent cells (Stratagene, La Jolla, CA, USA). Eighty-four positive clones were sequenced using the universal T7 primer and a BigDye v3.1 terminator kit on an ABI3730 DNA Analyzer automated sequencer (Applied Biosystems, Foster City, CA, USA).

The selection of sequences containing microsatellites was performed using WebSat (Martins et al., 2009). We considered dinucleotides with more than 6 repeats, trinucleotides, tetranucleotides, and pentanucleotides with 3 or more repeats. Primers were designed using PRIMER 3 (Rozen and Skaletsky, 2000). The software was configured to obtain primers with final amplification products ranging from 100 to 250 base pairs (bp) and primers ranging in size from 18 to 22 bp. Furthermore, the GC content should be 40-60%. To evaluate the quality of the initiators for the formation of secondary structures and to avoid the design of redundant primers, we used the Gene Runner 3.05 (http://www.generunner.net/) and Chromas 2 (http://technelysium.com.au/) software.

Microsatellite fragments were amplified using the thermocycler MyCycler Thermal Cycler (Bio-Rad, Hercules, CA, USA) in a total reaction volume of 16 μL, containing 20 ng genomic DNA template, 1 U *Taq* DNA polymerase (Fermentas, Vilnius, Lithuania), 10X polymerase chain reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.08% Nonidet P40), 0.32 mM of each dNTP, 2.4 mM $MgCl_2$, and 0.32 μM of each forward and reverse primers. Polymerase chain reaction was carried out in a touchdown cycling program (Don et al., 2001) under the following conditions: 94°C for 5 min, followed by 10 cycles at 95°C for 30 s, decreasing temperatures from 60° to 50°C by 1°C every cycle for 30 s, and 72°C for 50s followed by 30 cycles at 95°C for 30 s, 50°C for 30 s, and 72°C for 50 s; and a final extension at 72°C for 5 min. Amplification products were separated on denaturing sequencing gels (7% polyacrylamide and 7 M urea) in 1X TBE buffer (89 mM Tris base, 89 mM boric acid and 2 mM EDTA) at 60 W for 2 h. Gels were stained using a silver staining procedure (Creste et al., 2001) and photo-documented using a digital camera. The number of alleles and observed and expected heterozygosities were calculated using the R package (Jombart and Ahmed, 2011).

RESULTS AND DISCUSSION

We obtained and sequenced 84 colonies. Microsatellites were found in 57 colonies, representing 67.9% enrichment. Similar results were found for *Pimenta pseudocaryophyllus* (Myrtaceae), which showed 65.9% enrichment (Morgante et al., 2012). The values obtained in the present study were higher than those found for some other species; for example, 50.4% in *Aquilaria malaccensis*, a tropical climate species (Tnah et al., 2012); 40.3% enrichment obtained in *Mussaenda pubescens*, a shrub of the Rubiaceae family (Duan et al., 2012); and 35.9% in *Protium subserratum*, a species native to the Amazonian region (Misiewicz et al., 2012), such as annatto. However, a much higher value, 92.9%, was obtained for castor oil plant (*Ricinus communis*) (Bajay et al., 2009). Seventy SSRs were found in the 57 sequenced clones. Of these, 31 regions with sufficient quality were selected for primer design. The SSR markers had an average GC content of 47.3%, ranging from 40 to 57.9%, amplifying products with an average size of 186 bp, ranging from 122 to 228 bp, and average annealing temperature of 59.1°C, ranging from 56° to 62°C.

Of the 31 designed primers, 25 amplified consistent fragments (Table 1). Of these, 10 loci were polymorphic in the group of accessions studied, while 15 (60%) were monomorphic. However, in other groups of annatto accessions, these markers showed some polymorphism.

This low level of polymorphism (40%) was also obtained by other authors working with *Bixa orellana*. Carvalho et al. (2005) evaluated 60 genotypes annatto maintained in germplasm collections from various regions of Brazil and concluded that only 3 of the 21 isozyme loci were polymorphic (14.3%).

Table 1. Characteristics of 25 microsatellite loci of *Bixa orellana*.

SSR loci	GenBank accession	Repeat motif	Primer sequence (5'-3')	Ta (ºC)	Allele size (bp)
BorA2	KC152850	$(AC)_7$	F: CGAGGGGATGTGAAATTAGG R: CTGTTAGGGGACATTTTCATTG	TD_3	228
BorA7	KC152851	$(TCT)_5$	F: CTTTCCGCTTTTGGTCTTTG R: TGCAACCTTAATCCCTAGCA	TD_2	184
BorA9	KC152852	$(CAA)_5$	F: CTCCTTCCGCCAAAATCTC R: GTCCAAGATGCAGCAAATCA	TD_2	228
BorB4	KC152853	$(GA)_{16}$	F: GCATCATCTCTGTGGGATCA R: ACGAGAAAATGAACGAGCTG	TD_2	218
BorB6	KC152854	$(AC)_8$	F: TTCCCTCAGTTCTTCTTGAACC R: CGGAGCACTTTTCTCTTTGAG	TD_2	180
BorB10	KC152855	$(AG)_{14}$	F: CATGCCTCCTCTCGATCTCT R: ATCATTTCCTTGCCCAACTG	56	178
BorB12	KC152856	$(TG)_{14}(AG)_{19}$	F: CACATGCCCTTTGATGGTTA R: TCATGTTCCGATCCTATGTGA	56	172
BorC8	KC152857	$(CTT)_3$	F:CATTGTCAAAACCAGGAGGA R:GCTAGATTGGTGGCAAAAGG	TD_1	214
BorC11	KC152858	$(GAAA)_3$	F:TGCTCCATGAACTCGTCATT R:TGTGTGGGAGTTGGAAAACA	TD_1	181
BorC12	KC152859	$(CAT)_4$	F: GAGATTGTCGTGCATGGGTA R: AGGAGGAAGGAAAGGAGGTG	TD_2	194
BorD4	KC152860	$(AC)_7$	F: TCAAACATCGGCACATTGAT R: CCCCCTTTTGAATTGCTGT	TD_2	167
BorD7	KC152861	$(GT)_9$	F: GTTCGTGGTTGGAGCATACA R: CAGCAGGAACTGGAGGTGA	TD_2	209
BorD8	KC152862	$(AC)_8$	F: TGATGTGAAGAACCGAAATCC R: CAAAGCAGCAACCTCTTTTTC	TD_2	169
BorD12	KC152863	$(TG)_7$	F: GGATTCCATACATTTCAAGTCG R: CAGCATGATTGGTCAAAACTG	TD_2	207
BorE7	KC152864	$(AG)_5$	F: TCTCTTATCCCAGCGAGGAA R: CACATTTCTCTCTCCATGCA	TD_1	206
BorE12	KC152865	$(ATC)_3$	F: GTCATGCCAAGTTGTTTTTCC R: AGGTGTTTAGGCGAAAGTGG	TD_1	191
BorF1	KC152866	$(TG)_6(CT)_{17}$	F: CGTTCGTCCTGAAAATCTG R: CATTTCCCAGTGCAAGACC	TD_2	154
BorF4	KC152867	$(CTT)_3$	F: CTTCCTTGATTGCCTTTATTG R: TGCAGACCATTTGGAGTGA	TD_1	142
BorF5_2	KC152868	$(CA)_6$	F: GCGTGGACTAGCACTTTCTC R: GGGTGTGTTAGCCGAATTGTA	TD_3	168
BorF8	KC152869	$(AG)_{15}$	F: CACGGGGACAATTAAAGGAA R: TCGTTGATTGATGAGGATGC	TD_2	182
BorF9	KC152870	$(TATT)_3$	F: TCAACACACCACCACAATGA R: AGCTATGCCCAGAAAAAGGA	TD_1	122
BorG3	KC152871	$(GT)_7$	F: GTGTCACCCACAGAACACAA R: TGTTTCCAAGAAGCTAATGCAG	TD_2	175
BorG4	KC152872	$(CA)_{16}$	F: TCCCTTTGTTTTCCATTGCT R: TATCCACCAGCAACATTGAC	TD_1	209
BorG6	KC152873	$(GT)_8$	F: CGTGGACTAACAAGGACTGT R: TAATTTCTGCTTGGCACTCC	TD_1	130
BorH6	KC152874	$(AC)_{10}$	F: CAAGCACCCTTCTCTTTTCC R: GCTAAGGTTGATGGCAAATG	TD_2	160

Values based on 50 accessions of *Bixa orellana* germplasms of the Instituto Agronômico (IAC). Ta = annealing temperature; TD_1 = Touchdown PCR program with a temperature from 55°C, decreasing by 1°C per cycle until 45°C; TD_2 = Touchdown PCR program with a temperature from 60°C, decreasing by 1°C per cycle until 50°C; TD_3 = Touchdown PCR program with a temperature from 65°C, decreasing by 1°C per cycle until 55°C.

The 50 accessions analyzed with the 10 polymorphic loci showed 38 alleles, ranging from 2 to 6 alleles per locus, yielding an average of 3.8 alleles per locus. Values of expected heterozygosity ranged from 0.422 to 0.787, averaging 0.639. Moreover, the observed heterozygosity ranged from 0 to 0.658, averaging 0.541 (Table 2). Eight loci departed from Hardy-Weinberg equilibrium ($P < 0.05$). However, these results are expected when a germplasm collection is analyzed. In germplasm collections, genotypes are collected in different populations and did not breed since they were collected. Thus, Hardy-Weinberg equilibrium was not expected.

Our study results showed that the 10 microsatellite loci may be useful for evaluating genetic variation and population structure in *B. orellana* and will be used to assess local varieties and wild populations collected in the Amazon region, to further elucidate the genetic diversity and determine the distribution of this species in its center of origin and diversity.

Table 2. Alleles size in base pairs (A), expected heterozygosity (H_E), observed heterozygosity (H_O), Wright's fixation index (*f*) and P value for Hardy-Weinberg equilibrium (HWE) for 10 SSR loci developed for *Bixa orellana*.

SSR loci	Allele size (bp)	A	H_E	H_O	*f*	P value HWE
BorA2	228	3	0.537	0.200	0.628	0.000*
BorB4	218	3	0.422	0.182	0.569	0.000*
BorB10	178	6	0.787	0.489	0.379	0.000*
BorB12	172	5	0.629	0.208	0.669	0.000*
BorC12	194	3	0.488	0.204	0.582	0.000*
BorE7	206	3	0.588	0.000	1.000	0.000*
BorF1	154	4	0.674	0.440	0.347	0.000*
BorF5_2	168	5	0.527	0.658	-0.249	1.000
BorG3	175	2	0.453	0.286	0.369	0.009*
BorG4	209	4	0.519	0.479	0.077	0.397
Mean	-	3.8	0.639	0.541	0.437	-

*Significant at $P < 0.05$.

ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Processes #2011/05243-3, #2012/08307-5, and #2013/08884-5). We thank CNPq for providing scholarships to 3 of the authors.

REFERENCES

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

Arce J (1999). El Achiote *Bixa orellana* L. Cultivo Promisorio para el Erópico. 1st edn. Earth, Turrialba, Costa Rica.

Bajay MM, Pinheiro JP, Batista CEA, Nobrega MBM, et al. (2009). Development and characterization of microsatellite markers for castor (*Ricinus communis* L.), an important oleaginous species for biodiesel production. *Conserv. Genet. Resour.* 1: 237-239.

Billotte 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.

Carvalho JFRP, Robinson IP and Alfenas AC (2005). Isozymic variability in a Brazilian collection of annatto (*Bixa orellana* L.). *Pesq. Agropec. Bras.* 40: 653-660.

Carvalho PRN (1992). Corantes de urucum hidrossolúveis. *Rev. Bras. Corantes Naturais* 1: 242-243.

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

Creste S, Tulmann Neto A and Figueira A (2001). Detection of single sequence repeat polymorphisms in denaturing

polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Rep.* 19: 299-306.
Don RH, Cox PT, Wainwright BJ, Baker K, et al. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19: 4008.
Doyle JJ and Doyle JL (1990). Isolation of plant DNA fresh tissue. *Focus* 12: 13-15.
Duan TT, Gong W and Zhang DX (2012). Development of microsatellite markers from *Mussaenda pubescens* (Rubiaceae). *Am. J. Bot.* 99: e437-e439.
Fay MF, Cowan RS and Leitch IJ (2005). The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann. Bot.* 95: 237-246.
IBGE - Instituto Brasileiro de Geografia e Estatística (2009). Produção Agrícola Municipal. Available at [www.ibge.gov.br/estadosat/temas.php?sigla=df8tema=lavourapermanente2011]. Accessed April 15, 2014.
Jombart T and Ahmed I (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27: 3070-3071.
Kalia RK, Rai MK, Kalia S, Singh R, et al. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177: 309-334.
Lorenzi H (2002). Plantas Medicinais no Brasil: Nativas e Exóticas Cultivadas. Instituto Plantarum, Nova Odessa, São Paulo.
Martins WS, Lucas DC, Neves KF and Bertioli DJ (2009). WebSat-A web software for microsatellite marker development. *Bioinformation* 3: 282-283.
Medina A, Claret M, Ramis C and Diaz A (2011). Caracterización morfológica de frutos de onoto (*Bixa orellana* L.) y su correspondencia con patrons de proteinas e isozimas. *Acta Cient. Venezolana* 52: 14-23.
Misiewicz TM, Barbosa CE and Fine PV (2012). Microsatellite primers for an Amazonian lowland tropical tree, *Protium subserratum* (Burseraceae). *Am. J. Bot.* 99: e465-e467.
Morgante PG, Sebastiao I, Silveira LE, Mori GM, et al. (2012). Development of microsatellite markers for *Pimenta pseudocaryophyllus* (Myrtaceae), a wild South American species. *Am. J. Bot.* 99: e434-e436.
Rozen S and Skaletsky HJ (2009). PRIMER3. Available at [http://bioinfo.ut.ee/primer3-0.4.0/]. Accessed September 26, 2009.
Russell KR, Morrison EY and Ragoobirsingh D (2005). The effect of annatto on insulin binding properties in the dog. *Phytother. Res.* 19: 433-436.
Tnah LH, Lee CT, Lee SL, Ng KK, et al. (2012). Isolation and characterization of microsatellite markers for an important tropical tree, *Aquilaria malaccensis* (Thymelaeaceae). *Am. J. Bot.* 99: e431-e433.
Valdez-Ojeda R, Quiros CF, Aguilar-Espinosa ML and Rivera-Madrid R (2010). Outcrossing rates in annatto determined by sequence-related amplified polymorphism. *Agron. J.* 102: 1340-1345