University of São Paulo
"Luiz de Queiroz" College of Agriculture

Genetic structure, mating system and domestication of annatto (*Bixa orellana* L.)
using molecular markers

Gabriel Dequigiovanni

Thesis presented to obtain the degree of Doctor in
Science. Area: Genetics and Plant Breeding

Piracicaba
2017

Gabriel Dequigiovanni
Bioprocess and Biotechnology Engineering

Genetic structure, mating system and domestication of annatto (*Bixa orellana* L.) using
molecular markers

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **ELIZABETH ANN VEASEY**

Thesis presented to obtain the degree of Doctor in
Science. Area: Genetics and Plant Breeding

Piracicaba
2017

To those who I love without measures or proportions,
I just love them with all my heart.
My parents Alda and Enio,
My brother Diego,
My wife Patricia.

ACKNOWLEDGEMENTS

To the University of São Paulo, Superior School of Agriculture “Luiz de Queiroz” – ESALQ, and the Genetics and Plant Breeding graduate program for the opportunity to obtain the PhD degree.

To the São Paulo Research Foundation (FAPESP) for the financial support: grants 2013/08884-5 and 2016/05912-6, that were fundamental for the accomplishment of this work.

To the National Council for Scientific and Technological Development (CNPq) for the scholarship and financial support.

To Dr. Elizabeth Ann Veasey, my advisor and leader of the Plant Ecological Genetics Laboratory, for the confidence, all opportunities, discussions and the great help in the writing of this work and related projects. For never measure efforts in accomplishing this project, for the excellent corrections, and for the total support in all my decisions. I also thank you for being always on my side when the difficulties arose. To you Betty, my most sincere thanks.

To my dear parents, Enio and Alda, for the values, principles and unconditional love. Without the inspiration, guidance, and support that you have given me, I might not be the person I am today. I certainly owe everything to you.

To my brother Diego and my sister-in-law Carla, for the friendship, and even far away always sent many messages of encouragement.

To Schaker family, Jamil, Soili, Felipe and Leticia, for the support during all these years.

To Dr. Charles Roland Clement (INPA) for providing samples, and all the assistance in the delineation of the studies, in the corrections, in the articles writing, and especially for all the opportunities.

To Dr. Eliane Gomes Fabri (IAC) for helping in the selection and supply of plant materials, as well as the availability to solve doubts, and all her efforts that allowed me to carry out this work.

To my friends and colleagues of the Plant Ecological Genetics Laboratory: Jucelene, Mariana, Sueme, Nancy, Santiago, Adriana, Tatiane, Thais, Flaviane, Iris, Ana Flavia, Cattleya, Paula, Marcela, for the friendship and support during the development of this project.

To all the professors from the Genetics Department - ESALQ/USP, for the great teachings.

To the staff of the Genetics Department - ESALQ/USP, especially to Leia, for all the assistance.

To Marcos Cella (Marcão), technician of the Evolutionary Ecology and Applied Genetics Laboratory, for the help in the accomplishment of the experiments, and friendship.

To my special friends Sueme, Alessandro, Melina and Maisa, for making my days much happier.

To my friends Felipe and Juliana, who provided moments of great joy in this period.

To my friend Alessandro Alves Pereira for the great assistance in the collections and statistical analyzes, besides the discussions that have yielded results in this work and mainly for the friendship.

To my friend Santiago for the great help in the collections, laboratory analyzes, for the assistance in the period that I was in Manaus, and mainly for the friendship.

To Prof. Dr. Maria Teresa Gomes Lopes, for the opportunities, and especially for all assistance during the time I spent in Manaus.

To Prof. Dr. Maria Imaculada Zucchi, for the aid in the development of the enriched library and the great teachings.

To Prof. Dr. Marcos Siqueira, for the friendship and especially for opening the doors of ESALQ.

To Prof. Dr. Paul Gepts and his students for the great reception during my internship period at UCDavis, for the teachings and contributions to this work.

To Dr. Eduardo de Andrade Bressan for the suggestions during the development of the project, in writing and in the laboratory routines.

To Paulo and Marta from ITAL for sharing the phitochemical analyzes of the accessions from the germplasm bank;

To the “Orquestra Educacional de Piracicaba (OEP)” for the great moments we shared. Especially to maestro Ivan and my teachers Christine, Jacqueline and Marina.

To the technicians of EMATER-RO and EMATER-PA for assistance in the collections;

To all those who directly or indirectly contributed to the accomplishment of this work.

Especially to my wife Patricia, for her love, patience, understanding, support, encouragement and affection at all times. Without you nothing of this would be possible, thank you very much for everything.

*“If the sun refused to shine, I would still be loving you
When mountains crumble to the sea, there'll still be you and me...
And so today, my world it smiles; your hand in mine we walk the miles
Thanks to you it will be done. For you to me are the only one”*

J. P. Page and R. A. Plant (1969)

To God, for the gift of life, for always enlightening and guiding me.

*"Mama told me when I was young
Come sit beside me, my "little" son
And listen closely to what I say
And if you do this it will help you some sunny day*

*Oh, take your time, don't live too fast
Troubles will come, and they will pass
Follow your heart, and nothing else
And be a simple kind of man
be something you love and understand*

And don't forget, son there is someone up above..."

Lynyrd Skynyrd (1973)

RESUMO

Estrutura genética, sistema reprodutivo e domesticação de urucum (*Bixa orellana* L.) utilizando marcadores moleculares

Domesticação de plantas é um processo evolutivo que pode gerar uma série de modificações nas plantas para aumentar a adaptação para o cultivo e utilização pelos humanos. Estas modificações podem diminuir a aptidão das plantas no habitat selvagem, porém, aumentando sua aptidão para exploração humana. Urucum (*Bixa orellana* L.) é uma planta arbustiva domesticada na Amazônia a partir de populações de *Bixa orellana* var. *urucurana*. Esta tese apresenta um entendimento mais aprofundado sobre a domesticação, sistema reprodutivo e diversidade genética e estrutura de urucum e seu ancestral selvagem no Brasil. No primeiro estudo, um novo conjunto de 32 locos microssatélites foram isolados a partir de uma biblioteca genômica enriquecida com microssatélites, dos quais 12 foram polimórficos em populações de urucum selvagem e cultivado. No segundo estudo, a diversidade e estrutura genética de populações selvagens de urucum na Amazônia brasileira foram caracterizadas usando 16 marcadores microssatélites. Elevada estrutura populacional, e correlações positivas entre distâncias genéticas e geográficas foram observadas, sugerindo que a diferenciação genética é resultante de isolamento geográfico. Adicionalmente, Modelagem de Nicho Ecológico foi utilizada para caracterizar a distribuição potencial desta variedade no norte da América do Sul e observamos que o Sul de Rondônia, a bacia do rio Madre de Dios, os Llanos de Mojos e de Orinoco e oeste do Equador são áreas de alta probabilidade de ocorrência de urucum selvagem, fornecendo informações importantes para novas amostragens e conservação. No terceiro estudo, 16 locos de microssatélites e quatro compostos fitoquímicos foram utilizados para avaliar a diversidade genética de 63 acessos do banco de germoplasma de urucum do Instituto Agrônomo (IAC). Em ambas as análises, houve uma tendência de separação dos acessos de Rondônia, norte do Brasil, dos acessos do Sudeste. Os acessos de Rondônia apresentaram elevados valores para todos os compostos fitoquímicos e também apresentaram altos níveis de diversidade genética. Alguns acessos apresentaram níveis de bixina acima da média e são considerados materiais promissores para uso em programas de melhoramento genético de urucum. No quarto estudo, 12 locos microssatélites foram utilizados para determinar o sistema de cruzamento de uma população de urucum de Rondon do Pará, PA. A taxa de cruzamento multilocos indicou um sistema misto de cruzamento para esta população. A endogamia biparental também contribuiu para a taxa de autofecundação. Cruzamentos entre indivíduos aparentados também foram observados. Devido ao sistema misto, a coleta de sementes de polinização aberta para fins de conservação e melhoramento genético deve incluir pelo menos 60 plantas para assegurar uma amostragem representativa. No quinto estudo, a distribuição da diversidade genética entre amostras de urucum cultivado de quintais de comunidades ribeirinhas dos principais rios da Amazônia Brasileira, além de plantações ao longo das rodovias dos estados do Rondônia e Pará, além do Sudeste do Brasil foi caracterizada. As amostras coletadas apresentaram moderados níveis de diversidade genética e moderados a altos níveis de fluxo gênico entre os grupos geográficos, principalmente devido ao intercâmbio de semente entre agricultores. Contudo, análises Bayesianas e de agrupamento indicaram uma tendência de agrupamento baseado na origem geográfica das amostras. Isolamento por distância também foi observado de acordo com o teste de Mantel. No último estudo, amostras de urucum selvagem e cultivado da Amazônia brasileira foram comparados utilizando 16 locos microssatélites e duas regiões de DNA cloroplastidial. Uma clara separação entre cultivados e selvagens, suportada por altos valores de F_{ST} em ambas as análises foi observado. Amostras selvagens apresentaram altas taxas de diversidade em relação aos cultivados, parcialmente por não sofrerem seleção antrópica como acontece nas variedades cultivadas. Os dados sugerem a existência de relações genéticas entre urucum selvagem e cultivado, indicado por moderados níveis de fluxo gênico. Os resultados também demonstraram a proximidade entre grupos de urucum selvagem e cultivados de Rondônia e da bacia do Rio Madeira. Esta proximidade fornece indícios que a domesticação de urucum iniciou nesta região a partir de *B. orellana* var. *urucurana*.

Palavras-chave: *Bixa orellana*; Microssatélites; Amazônia; Diversidade genética; Estrutura genética; Domesticação; Cloroplasto

ABSTRACT

Genetic structure, mating system and domestication of annatto (*Bixa orellana* L.) populations from Brazil using molecular markers

Plant domestication is an evolutionary process that leads to several modifications in plants to increase adaptation to cultivation and utilization by humans. These modifications may decrease the fitness of plants in the wild habitat but increase it for human exploitation. Annatto (*Bixa orellana* L.) is a shrubby plant domesticated in Amazonia from wild annatto (*Bixa orellana* var. *urucurana*) populations. This thesis presents a more in-depth understanding of the domestication, mating system and genetic diversity and structure of annatto and its wild ancestor in Brazil. In the first study, a new set of 32 microsatellite loci isolated from a microsatellite-enriched genomic library was developed, of which 12 were polymorphic in populations of both cultivated and wild annatto. In the second study, the genetic diversity and structure of wild annatto populations in Brazilian Amazonia were characterized with 16 microsatellite markers. High population structure and positive correlation between genetic and geographic distances were found, suggesting that genetic differentiation might be caused by geographic isolation. Additionally, Ecological Niche Modeling was used to characterize the potential geographical range of this variety in northern South America and detected that South Rondônia, Madre di Dios River basin, Llanos de Mojos, Llanos de Orinoco and eastern Ecuador are highly suitable areas for wild annatto to occur, providing additional targets for future exploration and conservation. In the third study, 16 microsatellite loci and four phytochemical compounds were used to evaluate the genetic diversity of 63 accessions from the annatto germplasm bank at the Agronomic Institute (IAC). In both molecular and phytochemical analysis the results tended to separate the accessions from Rondônia, northern Brazil, from the Southwestern accessions. 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. In the fourth study, 12 microsatellite loci were used to determine the mating system of a cultivated population of annatto from Rondon do Pará, PA. Multilocus outcrossing rate indicated a mixed mating system for this population. Biparental inbreeding also contributed to the selfing rate in this population. Crossings among related individuals were also observed. Due to this mixed breeding system, the collection of open-pollinated seeds for plant breeding or conservation purposes should include at least 60 plants to ensure a representative sample. In the fifth study, the amount and distribution of genetic diversity among samples of cultivated annatto from homegardens of riverside communities along the major rivers in Brazilian Amazonia, and from farmer's fields along highways, in the States of Rondônia and Pará, and Southeastern Brazil was characterized. The samples collected presented moderate levels of genetic diversity, and moderate to high levels of admixture between geographic groups, occurring mainly due to exchange of seeds among farmers. However, cluster and Bayesian analyses showed a tendency to group samples based on their geographic origin. Isolation by distance was observed, according to Mantel's test. In the last study, wild and cultivated annatto samples from Brazilian Amazonia were compared using 16 microsatellite loci and two cpDNA regions. A clear separation between wild and cultivated annatto, supported by high values of F_{ST} in both analyses was observed. Wild samples presented higher rates of diversity in relation to cultivated, partly because these populations did not suffer anthropic selection, as in the cultivated varieties. The data suggest the existence of genetic relationship between wild and cultivated annatto, indicated by moderate levels of gene flow. The results also showed the proximity between groups of cultivated and wild accessions from Rondônia and the Madeira River basin. This proximity provides indications that annatto started its domestication in this area from *B. orellana* var. *urucurana*.

Keywords: *Bixa Orellana*; SSR; cpDNA; Genetic diversity; Genetic structure; Domestication; Amazonia

1. INTRODUCTION

Domestication of plants and animals is one of the most important processes taking place in the last 13,000 years, becoming a fundamental aspect in the transition from hunting-gathering to agriculture (Badr and El-Shazly, 2012). It was also a prerequisite to the rise of civilizations, and had transformed the global demography (Diamond, 2002). Plant domestication is the result of a selection process that leads to increased adaptation of plants to cultivation and utilization by humans (Brown, 2010). It is a continuing process over a long period of time, leading to significant changes in genotype and phenotype of the plants that are subjected to this process (Gepts, 2004). Plant domestication encompasses many evolutionary modifications that may decrease the fitness of a plant in wild habitat but increase it for human exploitation. The moment in which a crop can be considered domesticated is somewhat speculative. However, the full dependence of a plant on humans for survival is the fullest extent of domestication (Meyer and Purugganan, 2013).

The distinction between wild and cultivated is one of the basic questions of plant domestication (Terrell *et al.*, 2003). Domesticated plants present a number of morphological changes when compared to their wild ancestors. Harlan (1992) called these changes as "Domestication Syndrome". The traits in this syndrome include those increasing adaptation to cultivation and desirability of human consumption and use (Harlan, 1992; Gepts, 2004). Traits associated with domestication arise, at least in part, from human selection and thereafter relate to ways that plants are cultivated and harvested. Gepts (2004) also indicated that a common feature among most of the domesticated plants is a marked genetic bottleneck.

Amazonia is the world's most important center of biodiversity (Clement, 1989). At the time of European conquest, hundreds of species were domesticated in Amazonia, including cassava, cacao, sweet potato, pineapple and numerous fruit trees and palms. The ample number of crops domesticated in this area, besides cultural diversity, make Amazonia a very interesting area to study the domestication of plants (Diamond, 2002; Clement *et al.*, 2010, 2016).

Annatto (*Bixa orellana* L.) is a crop domesticated in Amazonia (Clement *et al.*, 2016). Moreira *et al.* (2015) proposed that *B. urucurana* is the wild ancestor of cultivated annatto, *B. orellana*. They also accepted Pilger's proposal, published by Kuntz, that *urucurana* is a variety of *B. orellana* (*B. orellana* var. *urucurana* (Willd.) Kuntze ex Pilg.). Our observations in the field, combined with information obtained from farmers, indicate that wild and domesticated types cross and generate fertile offspring. According to the biological concept of

species, this is sufficient to determine that they are not two separate species. Therefore, from the time of Moreira *et al.* (2015) publication, we are now denominating wild annatto as *B. orellana* var. *urucurana*.

In this study, several collections were made in Amazonia and during the expeditions we identified both cultivated and wild annatto populations and observed a series of domestication syndromes. Cultivated annatto presents large fruits, a large quantity of seeds per fruit and a very strong coloring of the seeds, besides showing great variability in fruit colors and shapes (Figure 1.1). On the other hand, the wild type presents small fruits, with few seeds and with little coloration in the aril of the seed, besides we observed only fruits of green color (Figure 1.2).

Annatto was domesticated by humans due to the dye that exists in the aril. This dye was widely used since ancient times. The Indians of tropical America made use of the colorful arils for body painting, dyeing of clothes, and especially for rituals (Plotkin, 1993). After the Europeans conquest, annatto became visible since Francisco de Orellana referred to annatto in one of his letters. The species was even named in honor of this navigator (Morton, 1960). Annatto was then widely used in the food and cosmetics industries in replace of synthetic pigments. The consumption of annatto increased since the prohibition of the use of synthetic dyes in some countries (Giuliano *et al.*, 2003; Nisar *et al.*, 2015). Brazil is the leader in annatto production, although Peru, Kenya, Colombia and other countries also grow annatto commercially. In Brazil, the states of Rondônia and São Paulo are the largest producers (Fabri, 2015).

Considering the economic and cultural importance of annatto, besides the fact that Brazil hosts great diversity for this crop, this work aimed to develop new microsatellite markers and generate information about its diversity, genetic structure, domestication and the reproductive system using microsatellite and cpDNA markers, plus phytochemical traits, studying both cultivated (*Bixa orellana*) and wild annatto (*B. orellana* var. *urucurana*), aiming to help to guide future research on breeding programs, as well as plan for the *in situ* and *ex situ* strategies for conservation of annatto in Brazil.

The specific objectives of this project are described below separately for each study.

Figure 1.1 Morphological variation on flowers, fruits and plant architecture of cultivated annatto (*Bixa orellana* L.)

Figure 1.2 Morphological variation on fruits and plant architecture of wild annatto (*Bixa orellana* var. *urucurana*)

1.1. Structure of the thesis and objectives

This thesis was organized in six studies presented in scientific manuscript format written in the English language. The studies were organized as follows:

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

Objective: Develop a new set of microsatellite loci for *B. orellana* and its wild relative *B. orellana* var. *urucurana* as a tool to generate useful information for conservation strategies and population genetic studies.

This study was submitted and accepted for publication in the journal *Crop Breeding and Applied Biotechnology*.

“Highly structured genetic diversity of *Bixa orellana* var. *urucurana*, the wild ancestor of annatto, in Brazilian Amazonia”

Objective: Evaluate the genetic diversity and structure of wild populations using microsatellite markers and characterize the potential geographical range of this variety in northern South America using Ecoclimatic Niche Modeling (ENM).

This study was submitted to the journal *PlosOne*.

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

Objective: 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.

This chapter was submitted and accepted for publication in the journal *Genetic Resources and Crop Evolution*.

“Genetic diversity and distribution of annatto (*Bixa orellana* L.) from Amazonia and Southeast of Brazil”

Objective: Characterize the amount and distribution of genetic diversity among cultivated accessions of cultivated annatto originated from the Amazonian region and also from Southern Brazil, in the State of São Paulo, using highly polymorphic microsatellite markers.

“Mixed mating system of annatto (*Bixa orellana*) determined with microsatellite markers”

Objective: Investigate the mating system of *Bixa orellana* in order to identify the levels of cross-fertilization and/or self-fertilization, and to better understand the genetic structure of progenies in a population of cultivated annatto from Brazilian Amazonia, providing information for conducting and establishing conservation and breeding programs.

“Domestication of annatto from *Bixa orellana* var. *urucurana*: Evidences based on SSR markers and chloroplast DNA”

Objective: In this study we aimed to answer the following questions: a) what are the levels of genetic diversity and genetic structure among cultivated annatto, compared to wild annatto? b) is there any evidence of gene flow, and therefore, of hybridization between wild and cultivated annatto? c) by the evidences gathered in this study, is it possible to suggest the most likely origin of domestication of annatto in Brazilian Amazonia?

REFERENCES

- Badr, A. and El-Shazly, H.** (2012) Molecular approaches to origin, ancestry and domestication history of crop plants: Barley and clover as examples. *J. Genet. Eng. Biotechnol.* **10**, 1–12.
- Brown, A.H.D.** (2010) Variation under domestication in plants: 1859 and today. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2523–2530.
- Clement, C.R.** (1989) A center of crop genetic diversity in western Amazonia. *Bioscience* **39**, 624–631.
- Clement, C.R., Cristo-Araújo, M. de, D’Eeckenbrugge, G.C., Pereira, A.A. and Picanço-Rodrigues, D.** (2010) Origin and domestication of native Amazonian crops. *Diversity* **2**, 72–106.
- Clement, C.R., Rodrigues, D.P., Alves-Pereira, A., Mühlen, G.S., Cristo-Araújo, M. De, Moreira, P.A., Lins, J. and Reis, V.M.** (2016) Crop domestication in the upper Madeira River basin. *Bol. do Mus. Para. Emilio Goeldi de Ciencias Humanas* **11**, 193–205.
- Diamond, J.** (2002) Evolution, consequences and future of plant and animal domestication. *Nature* **418**, 700–707.
- Fabri, E.G.** (2015) Demanda por corantes naturais aquece mercado brasileiro de urucum. *Soc. Nac. Agric.*
- Gepts, P.** (2004) Crop domestication as a long term selection experiment, Oxford, UK: John Wiley & Sons, Inc.
- Giuliano, G., Rosati, C. and Bramley, P.M.** (2003) To dye or not to dye: biochemistry of annatto unveiled. *Trends Biotechnol.* **21**, 513–6.
- Harlan, J.** (1992) Crops and man, Madison: American Society of Agronomy.
- Meyer, R.S. and Purugganan, M.D.** (2013) Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14**, 840–852.
- Moreira, P.A., Lins, J., Dequigiovanni, G., Veasey, E.A. and Clement, C.R.** (2015) The domestication of Annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. *Econ. Bot.* **69**, 127–135.

- Morton, J.F.** (1960) Can Annato (*Bixa orellana* L.) an old source of food color, meet new needs for safe dye. *Florida State Hortic. Soc.*, 301–309.
- Nisar, N., Li, L., Lu, S., et al.** (2015) Carotenoid metabolism in plants. *Mol. Plant* **8**, 68–82.
- Plotkin, M.** (1993) *Tales of a Shaman's apprentice*, New York: Penguin Books.
- Terrell, J.E., Hart, J.P., Barut, S., et al.** (2003) Domesticated landscapes: the subsistence ecology of plant and animal domestication. *J. Archaeol. Method Theory* **10**, 323–368.