FEKADU GEBRETENSAY MENGISTU

CROSS-SPECIES AMPLIFICATION OF MICROSATELLITE MARKERS AND GENETIC DIVERSITY IN THE MACAW PALM (*Acrocomia aculeata)*

Thesis submitted to the Universidade Federal de Viçosa, as part of the requirements of the Post-Graduate Program in Plant Sciences, to obtain the title of *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRAZIL
2015

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
M544c
2015

Mengistu, Fekadu Gebretensay, 1978-
Cross-species amplification of microsatellite markers and genetic diversity in the macaw palm (*Acrocomia aculeata*) / Fekadu Gebretensay Mengistu. – Viçosa, MG, 2015.
xiv, 84f. : il. ; 29 cm.

Orientador: Sérgio Yoshimitsu Motoike.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Acrocomia aculeata*. 2. Macaúba - Melhoramento genético. 3. Marcadores moleculares. 4. Diversidade genética. 5. Germoplasma vegetal. 6. Biocombustível. I. Universidade Federal de Viçosa. Departamento de Fitotecnia. Programa de Pós-graduação em Fitotecnia. II. Título.

CDD 22. ed. 633.851

FEKADU GEBRETENSAY MENGISTU

CROSS-SPECIES AMPLIFICATION OF MICROSATELLITE MARKERS AND GENETIC DIVERSITY IN THE MACAW PALM (*Acrocomia aculeata*)

Thesis submitted to the Universidade Federal de Viçosa, as part of the requirements of the Post-Graduate Program in Plant Sciences, to obtain the title of *Doctor Scientiae*.

APPROVED: July 28, 2015.

Kacilda Naomi Kuki

Bruno Galvêas Laviola

Eveline Teixeira Caixeta
(Co-adviser)

Cosme Damião Cruz
(Co-adviser)

Sérgio Yoshimitsu Motoike
(Adviser)

Dedicated to my family especially my mother Mrs. Aregash Admassie; my wife Etsegenet Tsegaye and my children Hermella and Amanuel Fekadu

ACKNOWLDEGEMENT

My first most special thanks go to the almighty God!

I would like to acknowledge the Postgraduate Program of Plant Sciences, Federal University of Viçosa, for accepting my application to study doctor science.

The professors of the Postgraduate Program of Plant Sciences and Genetics and Breeding are also acknowledged for the knowledge they shared with me and support for the last four years.

Special thanks also go to my beloved wife Etsegenet Tsegaye; my children Hermella and Amanuel Fekadu; my mother Aregash Admassie and the rest of family members for their unreserved support and endurance during this long period of my absence away from them.

I would like to thank Dr. Gustavo Fialho; Dr. Fabricio Sobreira; Prof. Cláudio Horst Bruckner and for their special reception and treatment from the first day of my arrival in Brazil. I am also grateful to my advisor Prof. Sérgio Yoshimitsu Motoike for his personal encouragement, research advisor-ship and professional guidance until this moment. My special thanks also go to colleagues of Biotechnology and Plant Breeding Laboratory, Department of Plant Sciences, including Dra. Telma, Dr. Éder Lanes, Prof. Carlos Nick, Renata Freitas, Dra. Kacilda Naomi Kuki and Fransico for their unreserved assistance during my laboratory and field activities.

Petróleo Brasileiro S.A (Petrobras); Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and The Wold Academy of Sciences (TWAS) are highly acknowledged for their financial support.

Finally, I am grateful to my roommates Bayisa Asefa and Wogayehu Worku for their love, support and good times I passed with them. Dr. Alemayehu Mideksa, from University of California, USA is also acknowledged for his technical help to develop aerial map of data points in the thesis work.

BIOGRAPHY

Fekadu Gebretensay Mengistu was born in 1978 in East Showa, Debrezeit town about 50 km away from the capital Addis Ababa, Ethiopia. He attended his elementary education at Keta Primary School from 1983 to 1991 and high school education in Harer Meda Model Senior Secondary School in Debrezeit from 1992 to 1996. He joined Awassa College of Agriculture in 1996 and graduated in 2000 with B.Sc. degree in Plant Production and Dry Land Farming. After graduation, he worked in the Ministry of Agriculture (MoA), Woreilu Woreda Agricultural Office in West Wello as agronomist for one year from 2000 to 2001. He then employed by the late Ethiopian Agricultural Research Organization (EARO), currently Ethiopian Institute of Agricultural Research (EIAR) in March 2001 and worked as researcher in horticulture research section for 5 years at Kulumsa Agricultural Research Center (KARC), one of the Federal Agricultural Research Centers under EIAR. Then he joined his M.Sc. program in University of Copenhagen, Faculty of Life Sciences in Denmark in 2007 and studied for 2 years and graduated in Horticulture Science in 2009. He then re-employed by EIAR/KARC on the same post as Associate Researcher and worked until 2011. Finally, he was accepted as a postgraduate student in the Federal University of Viçosa, Postgraduate Program of Plant Sciences in August 2011. He finally submitted his doctor thesis to defend in July 2015.

TABLE OF CONTENTS

LIST OF SYMBOLES AND ABBREVIATIONS

A	Number of alleles per locus
F	Inbreeding coefficient
H_E	Expected heterozygosity
H_O	Observed heterozygosity
$\bar{H}_O$	Average observed heterozygosity
HWE	Hardy-Weinberg Equilibrium
I	Nei's genetic identity
N_a	Average number of alleles per provenance
N_e	Effective number of alleles per provenance
N_t	Total number of alleles per pronenance
P	Percentage of polymorphic loci
p	Frequency of null alleles
P_a	Proportion of alleles per provenance
Φ	Phi-statistics
$\hat{\sigma}_a^2$	Variance among provenances
$\hat{\sigma}_b^2$	Variance among families
$\hat{\sigma}_c^2$	Variance within family
$\hat{\sigma}_T^2$	Total variance
$AgNO_3$	Silver Nitrate
AMOVA	Analysis of Molecular Variance
BAG	Active Germplasm Bank of Macaúba
CTAB	Cetyl Tri-methyl Amonium Bromide
dNTPs	Deoxynucleotide triphosphates
EDTA	Ethylene Diamine Tetra Acetic Acid

gDNA	Genomic DNA
HCL	Hydrochloric Acid
$MgCl_2$	Magnesium Chloride
NaCl	Sodium Chloride
Na_2CO_3	Sodium carbonate
PCoA	Principal Coordinate Analysis
PCR	Polymersase Chain Reaction
PIC	Polymorphic Information Content
SDS	Sodium Dodecyl Sulfate
SSRs	Simple Squence Repeats
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
UPGMA	Un-weighted Pair Group Arthemetic Mean Average

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF TABLES

CHAPTER 1

CHAPTER 2

RESUMO

MENGISTU, Fekadu Gebretensay, D.Sc., Universidade Federal de Viçosa, julho de 2015. **Amplificação cruzada de marcadores microssatélites e estudo da diversidade genética em palmeira macaúba (*Acrocomia aculeata*).** Orientador: Sérgio Yoshimitsu Motoike. Coorientadores: Eveline Teixeira Caixeta e Cosme Damião Cruz.

A palmeira macaúba (*Acrocomia aculeata*) é espécies oleaginosas na América do Sul com abundante distribuição natural no Brasil. A macaúba é considerado como um grande potencial para a produção de biodiesel e estar sob domesticação no Brasil. Recentemente pesquisas mostram que macaúba está sob a ameaça de extrativismo predatório, as alterações climáticas e as políticas de uso da terra na população natural e precisa ser conservada *ex situ* para o futuro melhoramento genético e uso sustentável de seus recursos genéticos. Os microssatélites (Simple Sequence Repeats-SSR) são um dos marcadores moleculares mais aplicáveis na caracterização de coleções de germoplasma e ajudar na conservação da variabilidade genética em bancos de germoplasma. No passado, apenas alguns marcadores SSR foram desenvolvidos para a macaúba, devido ao alto custo do desenvolvimento e da limitação do conhecimento sobre o potencial da macaúba. Dois experimentos foram conduzidos neste estudo: (1) para demonstrar o uso de amplificação cruzada de marcadores SSR como uma alternativa de baixo custo para estabelecer os marcadores SSR para *A. aculeata*; e (2) para estudar a diversidade genética nas coleções *ex situ* de germoplasma da *A. aculeata* que foram originalmente coletadas de diferentes procedências no Brasil. Na primeira parte do trabalho, um estudo de amplificação cruzada realizado para avaliar a possibilidade de transferência de 34 marcadores SSR, originalmente desenvolvidos para duas espécies de Arecaceae (*Astrocaryum aculeatum* e *Elaies oleifera*) em *A. aculeata* usando 192 acessos de 41 famílias originalmente oriundos de seis procediências no Brasil. Do total de marcadores avaliados, 15 SSR (44%) amplificaram com sucesso o DNA genômico de *A. aculeata*, dos quais quatro SSR (26%) foram polimórficos. O baixo sucesso do amplificção cruzada foi devido á relativa distância taxonómica entre as fontes (*A. aculeatum* e *E. oleifera*) e as espécies-alvo (*A. aculeata*). No entanto, os marcadores polimórficos identificados pela transferência detectaram uma média elevada de locos polimórficos (P = 79%) por procediência. Os marcadores também revelaram a deficiência de heterozigotos nos acessos analisados, e que for confirmado pelo coeficientes de endogamia positivos obtidos em todos os locos. No segundo trabalho, estudo da

diversidade genética foi realizada nos 192 acessos de *A. aculeata* com base em dez marcadores SSR (incluindo dois SSR polimórficos identificados no estudo da transferibilidade e o resto de oito SSR foram seleccionados a partir de grupos de SSR anteriormente desenvolvidos para *A. aculeata*). O estudo resultou em diferentes níveis de diversidade alélica, heterozigosidade e polimorfismo entre os acessos analisados. Com base em distâncias genéticas, três grupos distintos de procediências foram estabelecidos utilizando diferentes métodos de agrupamento. No entanto, o teste de Mantel revelou uma correlação não significativa entre a distância genética e geográfica entre as procediências. A proporção da variação genética foi estimada pela análise de variância molecular (AMOVA), que revelou maior variação genética dentro da família do que entre as famílias de *A. aculeata,* seguido de variação entre as procediências. O estudo comprovou a eficiência de transferência inter-espécies de marcadores SSR entre diferentes espécies de Arecaceae. Ele também reafirmou que SSR são marcadores moleculares úteis na caracterização de germoplasma de *A. aculeata* e as informações que foram gerados podem ser utilizados em conservação de germoplasma e no programa de melhoramento da *A. aculeata*. Os marcadores podem ser utilizados para ajudar o programa de melhoramento genético através de genotipagem de cada indivíduo no banco de germoplasma que ajudaria a identificar grupos geneticamente distintos e também minimizar a redundância de entradas no banco de germoplasma que potencialmente maximizar a eficiência de conservação e reduzir seus custos.

ABSTRACT

MENGISTU, Fekadu Gebretensay, D.Sc., Universidade Federal de Viçosa, July, 2015. **Cross-species amplification of microsatellite markers and genetic diversity in the macaw palm (*Acrocomia aculeata*).** Adviser: Sérgio Yoshimitsu Motoike. Co-advisers: Eveline Teixeira Caixeta and Cosme Damião Cruz.

Macaw palm (*Acrocomia aculeata*) is newly emerging oleaginous species in South America with abundant natural distribution in Brazil. It is considered as a great potential palm for production of biodiesel and being under domestication in Brazil. Recently researches show that macaw palm is under the threat of predatory extractivism, climate change and land use policies in the natural population and need to be conserved *ex situ* for future genetic improvement and sustainable use of its genetic resources. Microsatellites (Simple Sequence Repeats-SSR) are one of the most applicable molecular markers in characterization of germplasm collections in plants and help to conserve genetic variability in germplasm banks. In the past only few SSR markers were developed for the macaw palm owing to the high development cost and limitation in knowledge about the importance of the palm. Two experiments were set up in this study: (1) to demonstrate the use of cross-species amplification as a cost-effective altarnative to establish SSR markers for *A. aculeata;* and (2) to study the genetic diversity in *A. aculeata ex situ* germplasm collections which were originally collected from different provenances in Brasil. In the the first part of the work, a cross-species amplification study was conducted to evaluate the transferability of 34 SSR markers, originally developed for two Arecaceae species (*Astrocaryum aculeatum* and *Elaies oleifera*) in *A. aculeata* using 192 accessions from 41 families collected from six provenances in Brazil. From the total markers evaluated, 15 SSR (44%) successfully amplified the genomic DNA in *A. aculeata*, of which four SSR (26%) were polymorphic. The low success of the cross-amplification was accounted for the relatively wider taxonomic distance between the sources (*A. aculeatum* and *E. oleifera*) and the target (*A. aculeata*) species. However, the polymorphic markers identified by this study detected a high average percentage of polymorphic loci (*P*=79%) per provenance. The markers also revealed heterozygote deficiency in the accessions and this was confirmed by positive inbreeding coefficients obtained in all the loci analyzed. In the second work, genetic diversity study was carried out in the 192 accessions of *A. aculeata,* based on ten SSR markers (including two polymorphic SSR indentified in the transferability study and the rest eight from sets of SSR previously developed for *A.*

aculeata). The study resulted in different levels of allelic diversity, heterozygosity and polymorphism among the accessions analyzed. Based on genetic distances, three distinct groups of provenances were established using different methods of grouping. However, Mantel test detected a non-significant correlation between the genetic and geographic distances among the provenances, revealing genetic similarities among geographically distant provenances in the country. Analysis of Molecular Variance (AMOVA) revealed the proportions of genetic variation, in which more genetic variation was obtained within family than among families of *A. aculeata* followed by variation among provenances. The study proved the efficiency of inter-species transferability of SSR markers between different species in Arecaceae. It also reaffirmed that SSR are useful molecular markers in characterizing *A. aculeata* germplasm and the information that were generated could be utilized in *A. aculeata* germplasm conservation and breeding program. The markers could be used to help the breeding program through genotyping each individuals in the germplasm bank that would help to identify genetically distinct groups and also minimize the unnecessary redundancies of entries in the germplasm bank that would potentially maximize conservation efficiencies and reduce its costs.

GENERAL INTRODUCTION

Macaúba or also known as Macaw palm (*Acrocomia aculeata* (Jack.) Lodd. Ex.Mart.) is a perennial arborescent palm tree, native to tropical forests, mainly in South America with wide geographical distribution in Brazil (Scariot et al., 1995). It stands out among the promising oil palms for biodiesel production. It has possibility of becoming commercially the most important oil palm in the Brazilian context, highlighting not only by its considerable oleaginous potential, but also for being species native to Brazil, which occur in large natural mass in much of the national territory (Silva, 1994). This specie has broad utility and more recently its fruit has aroused great social and economic interest in its vegetable oil production capacity, considering that macaúba is cited as one of the main sources of oil in Brazil (Nucci et al., 2008). The oil extracted from macaúba fruits has been used in cosmetic industries, medicinal uses, oil and flour used in food and, most recently as feedstock for biodiesel production (Bora and Rocha, 2004; Hiane et al., 2005).

However, the introduction of macaúba oils in biodiesel production still depends on the solution of some challenges. Among them, the development of commercial plantations, which highly relies on the availability of suitable varieties which could meet the required demands of producers and manufacturers; such as high fruit productivity, good oil quality, and other desirable agronomic characteristics. Moreover, the pressure from the local extractivists together with climate changes and land use policies would contribute for the degradation of the natural populations of this important palm species (Faleiro et al., 2008). Therefore, genetic resource conservation is one of the primary steps currently undertaken in the process of domesticating this noble palm species. The study of the genetic characterization from an *ex situ* collections in macaúba is unprecedented and is very important because it brings relevant subsidies for the identification and conservation of genetic variability, helping still in different

stages of pre-breeding of this palm and the bases for future development of suitable varieties. To quantify the genetic variability between and within populations, the use of molecular markers has proved to be important and efficient tools. For this study, microsatellite markers were chosen for their intrinsic characteristics of stable marker, co-dominant, multiallelic, robust, high reliability, the operational simplicity and speed of analysis when compared to other methods (Mohan et al., 1997; Agarwal et al., 2008).

Since macaúba is a newly emerging species, only few molecular markers such as microsatellites are currently available. Therefore, in our study, we did a cross-species transferability of some polymorphic microsatellite markers from related palm species in Arecaceae family (*Astrocaryum aculeatum* and *Elaies oleifera)* as an alternative to select some working markers for *A. aculeata*. Cross-species amplification and transferability studies reveal the capabilities of SSR markers to cross amplify homologous sites among related species and genera in Arecaceae. Ramos et al. (2012) and Zaki et al. (2012) reported cross-amplification and transferability of SSR markers from *A. aculeatum* and *E. oleifera,* respectively in several species of Arecaceae. *A. aculeatum* knows as Tucumã of Amazonas is a tropical palm found in western and central Brazilian Amazonia. The raw fruit is traditionally eaten by the local population and edible oil can be pressed from the fruits for human consumption as well as the dried cake residue is used as animal feed (Ramos et al., 2012). Whereas *E. oleifera* is a species in the oil palm genus along with the commercial *Elaeis guineensis* and occurs naturally in south-central America and in the Amazon (Teh, 2010). *E. oleifera* has been a promising genetic resource for oil palm improvement and currently used in oil palm hybrid breeding programs (Zaki et al., 2012).

This study had two main objectives: establishment and characterization of microsatellite markers for *A. aculeata* using cross-amplification and transferability from two related palm species in Arecaceae and study genetic diversity in *ex situ* germplasm

collections of *A. aculeata* using microsatellite markers including from the previsouly developed sets of SSR for macaúba.

Hence, the specific objectives were:

a) to evaluate the cross-amplification and transferability of some microsatellite markers sourced from Tucumã (*A. aculeatum*) and south America oil palm (*E. oleifera*), in accessions of *A. aculeata*;

b) to characterize the identified polymorphic microsatellite markers using different parameters in accessions of *A. aculeata;* and

c) to quantify the amount of genetic variability among and within *A. aculeata* germplasm collections in a germplasm bank.

REFERENCES

Agarwal M, Shivastava N and Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27: 617-631.

Bora PS and Rocha RVM (2004) Macaíba palm: fatty and amino acids composition of fruits. *Cienc. Tecnol. Aliment*. 4: 158–162.

Faleiro FG, Costa AM, Karia CT, Andrade RP et al. (2008) Molecular markers and geographic information systems as a tool to study native plant species in the Brasilian Savannas. Available at http://www.cpac.embrapa.br.

Hiane PA, Ramos-Filho, MM, Ramos MIL and Macedo MLR (2005) Bocaiúva, *Acrocomia aculeata* (Jacq.) Lodd., pulp and kernel oils: characterization and fatty acid composition. *Brazilian Journal of Food Technology* 8: 256–259.

Mohan M, Nair S, Bhagwat A, Krishna TG et al. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* 3: 87-103.

Nucci SM, Azevedo-Filho A, Colomo AC, Priolli GHR, Coelho MR, Mata TL and Zucchi IM (2008) Development and characterization of microsatellites markers from the macaw. *Molecular Ecology Resources* 8: 224–226.

Ramos SF, Vasconcelos de Macêdo JL, Lopes MTG, Batista JS et al. (2012) Microsatellite Loci for Tucumã of Amazonas (*Astrocaryum aculeatum*) and amplification in other Arecaceae. *American Journal of Botany* 508-510. doi:10.3732/ajb.1100607.

Scariot AO, Lleras E and Hay JD (1995) Flowering and fruiting phenologies of the palm *Acrocomia aculeata*: patterns and consequences. *Biotropica* 27:168-173.

Silva JC (1994) Macaúba: fonte de matéria-prima para os setores alimentício, energético e industrial. Trabalho de conclusão da disciplina Cultivo de essências exóticas e nativas. Departamento de Engenharia Florestal, Universidade Federal de Viçosa.

Teh CK (2010) Genetic Diversity of Central and South American Wild Oil Palm (*Elaies oleifera*) Populations Using Microsatellite Markers. M.Sc. Thesis. Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

Zaki NM, Singh R, Rosli R, Ismail I (2012) *Elaeis oleifera* Genomic-SSR Markers: Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. International Journal of Molecular Sciences 13: 4069-4088.

LITERATURE REVIEW

1. BOTANY AND TAXONOMICAL DESCRIPTION

Acrocomia aculeata has a wide popular synonym in Brazil: macaúba, mucajá, mocujá, mocajá, macaw, macaiúva, bacaiúva, bocaiúva, imbocaiá, coco-de-phlegm or coco-de-spinho (Lorenzi, 1992). It was formely under family Palmae, is currently classified within the Arecaceae family and genus Acrocomia (Henderson *et al*., 1995). Taxonomically, *A. aculeata* is classified under: Kingdom: Plantae; Division: Magnoliophyta; Class: Liliopsida; Subclass: Arecidae; Order: Arecales; Family: Arecaceae; Subfamily: Arecoideae; Tribe: Cocoeae; Subtribe: Bactridinae; Genre: *Acrocomia* and species: *Acrocomia aculeata* (Grin, 2013).

According to cytogenetic studies in fifty species of twenty eight genera of palms, including some species in the subfamily Arecoideae, different numbers of chromosomes was found (Sharma and Sarkar, 1956). However, there is a general homogeneity and uniformity of clearly marked chromosomes within each subfamily, and all species are diploid. They consider the subfamily Arecoideae as a good example of evolutionary stability, in which in six different species of different genres, the same number of chromosomes (2n= 2x = 32) and an overall similarity in karyotypes were found. However, Read (1966) did new chromosome count in Palmae (Arecaceae), which divides the subfamily Arecoideae into two species: species with n = 15, which include *A. aculeata* and species with n = 16.

The family Arecaceae which belongs to macaúba, is composed of a group of generically known species such as palm trees, with towering size that easily distinguishes it from other plants (Lorenzi *et al*., 1996). Although there are reports indicating the existence of several species under the genus Acrocomia, it consists of two main species, *A. aculeata* (Jacq.) Lodd ex Mart and *A. hassleri* (B. Rodr.) WJH ahn.

These species are mainly in tropical distribution, but have worldwide occurrence. The number of genera could reach to 236 and 3400 species (Aguiar, 1988; Lorenzi *et al*., 2004).

The term Acrocomia is Greece origin to describe the arrangement of the leaves in the plant. "Akron" (top) and "Kome" (hair) suggesting that the leaves are arranged in the shape of a crown (Novaes, 1952; Henderson *et al*., 1995). Due to the co-existence of one or more Acrocomia species, there is high variability of *A. aculeata* in natural stands as well as in germplasm banks. However, there still exists limitation of well established descriptors or phylogenetic reports available to distinguish between the species. However, efforts have been made to study the morphological structure of *A. aculeata*, which believe to be of a species widely and abundantly distributed in Brazil and becoming center of several researches due to its high fruit productivity and oil quality for biodiesel production. The morphological reports described here are based on some researches done in the past and could be used as reference to at least have some detail until the species is domesticated and detail reports are available.

A. aculeata is represented by arborescent palms and it is evergreen whose stem can reach up to 15 m in height with 20 to 30 cm in diameter. Most plants have long dark spines along the trunk and the base of the leaf sheaths. The leaves are pinnate presenting 4 to 5 m long, usually 20 to 30 in numbers, dark green color and contemporary distributed in different planes giving a feathery appearance to the crown. The leaflets are approximately 130 in number on each side of the central region of the leaf sheath and are covered with spines (Lorenzi, 1992).

The blossoms are yellow, arranged in small bunches at 80 cm long, protected by a spathe of varying size, which can reach up to 2 m in length. The flowers are small, unisexual, and both sexes are together in the same inflorescence, but the male flowers are at the top of the inflorescence (Silva *et al*., 2001). According to Scariot *et al*. (1991),

flowering in *A. aculeata* occurs between August and December and inflorescences are androgynous with marked protogyny. The principal pollinators are mainly beetles: *Andranthobius* sp. (Coleoptera: Curculionidae), *Mystrops cf Mexicana* (Nitidulidae) and *Cyclocephala forsteri* (Scarabaeidae), with wind playing a secondary role. Cross-pollination between different individuals (xenogamy) accounts for most of the reproductive system. However, the species is self-compatible with geitonogamy accounting for a significant percentage of fruit set (Scariot *et al*, 1991).

Each inflorescence produces an average of 60 fruits, but this number can vary from 0 to 271, according to the region (Scariot *et al*., 1995). Fruiting could occur throughout the year, especially from September to January (Scariot *et al*., 1991; Lorenzi, 1992; Lorenzi, 1996; Silva *et al*., 2001; Nucci, 2007). A macaw palm can produce 3 to 6 bunches per year with an average of 800 fruits per cluster, ensuring the dispersion of the plant. The fruit is composed of 42% pulp (mesocarp); 20% epicarp (shell) and the remainder is formed by diaspore of the fruits consisting of endocarp (31%) and seeds (7%). The fruits are spherical or slightly flattened, shaped in globose drupe with a diameter ranging from 2 to 5 cm, and form one to three seeds or embyos. The epicarp is thin and easily broken up when mature. The mesocarp is fibrous, mucilaginous, sweet tasting, rich in glycerides, yellow or yellow-brownish coloring (Chuba *et al*., 2008; Ramos *et al*., 2008).

2. GEOGRAPHICAL DISTRIBUTION AND ECOLOGY

According to Henderson *et al*. (1995), *A. aculeata* has large and wide distribution in the dry regions of tropical America (Figure 1). It is native palm tree of tropical forests, occurring from southern Mexico to southern Brazil, Paraguay and Argentina (Morcote-Rios and Bernal, 2001). In Brazil, it is abundantly found in the States of Ceará, Minas Gerais, Mato Grosso, Mato Grosso do Sul, São Paulo, Parana,

Santa Catarina and Rio Grande do Sul. Its distribution area has been heavily influenced by human activities. In Brazil, the palm greater dispersion was with the occurrence of natural stands of this species in almost every territory and widely spread and used by Cerrado people (Silva, 1994; Henderson *et al*., 1995).

Figure 1. Geographical distribution of *Acrocomia aculeata* in South America (Adopted from Nucci (2007); original source: New World Fruits (2014)).

According to Scariot *et al*. (1991, 1995), *A. aculeata* is cited as the single tree-sized species in pasture areas and occurs mainly in semi-deciduous broadleaved forests (Lorenzi, 1992). This palm has adaptations in different environmental conditions. It was described as mesotrophic soils indicator of central Brazil (Ratter *et al*., 1996). However, there are reports showing in the state of São Paulo, it mainly occurs in poor soils covered by savannah (Novaes, 1952).

3. AGRICULTURAL PRACTICES AND USES

A. aculeata has been identified suitable for Agricultural practices in tropics mainly in degraded lands, where resources are scarce to grow agricultural crops.

However, if managed properly, in arable lands, it can boost integral agricultural productivity as it can grow mutually with other crop plants and provides spaces for livestock grazing (Averdunk *et al*., 2013). *A. aculeata* is cited as one of main sources of vegetable oil in Brazil, which has the potential to produce biodiesel. Macaúba can yield up to 4000 liters of vegetable oil/ha/year, which is high as compared to that of commonly cultivated annual crops such as soybeans which produces only 420 liters, sunflower 890 liters and even castor only 1320 liter/ha/year (Cesar, 2003). Besides, from the solid waste, products such as charcoal and animal feeds could be obtained as by-products (Oliveira, 2006).

In addition the pulp of the fruit is consumed fresh or used for edible oil extraction, and products such as candy, and beverages are prepared out of it. Other parts of the palm such as straws to make crafts and construction of houses by the traditional people, the pulp and seeds are used as food and ingredients for paints and the trunks for their traditional logs. Besides, several species of wild animals feed on its leaves, fruits of pulp and seeds. The core of the stem gets a nutritious starch and the leaves and fodder provide textile fibers for making nets and fishing lines (Lorenzi, 1992).

Besides, *A. aculeata* has been identified as important palm tree from natural, economical and ecological point of view with the advantage to grow with little water and strive in drought prone and degraded areas (Lima *et al*., 2003). *A. aculeata* offers further advantages over agricultural crops that occupy today prominent position in Brazil in biodiesel production, such as soybean and palm oil. Hence, with regard to manpower, it creates more job opportunities which require manual work during harvest and it can provide family and small producers an alternative source of income (Holanda, 2004).

Owing to the widespread diversity and distribution of *A. aculeata* in Brazil and its immense potential for biodiesel production, studies related to conservation of the

genetic resources are vital. It has fundamental importance for sustainable exploitation and conservation of its genetic variability present in its germplasm collections, given the lack of studies on the species.

4. GENETIC DIVERSITY IN PLANTS

Genetic diversity is the measure of genetic variability within and among a given population through different ways such as identifying alleles, the number of individuals and type of genotypes being heterozygous as well as homozygous in different proportions or different amount (Weir, 1996). Agricultural scientists realized that plant genetic diversity can be captured and stored in the form of plant genetic resources such as gene bank, DNA library, and so forth, in the bio-repository which preserve genetic material for long period. However, conserved plant genetic resources must be utilized for crop improvement in order to meet future global challenges in relation to food and nutritional security (Govindaraj *et al*., 2015).

To quantify the genetic variability in plant populations, the most commonly used genetic parameters are the number of alleles per locus, the expected and observed heterozygosity (Pinto, 2001). The genetic alteration of a population can be assessed by the change in their gene frequencies, which makes an estimated frequency of heterozygotes in fundamental evolutionary studies (Nei, 1978). A specific allele in a set of diploid individuals is often twice the number of individuals homozygous for the allele, plus the number of individuals in the group. According to Weir (1996), the frequency of heterozygotes is important in that each heterozygous carries different alleles, demonstrating the existence of genetic variation in the population. Evolutionary forces such as mutation, migration, selection and genetic drift act within the context of each species influencing the genetic structure of the population, i.e. interfering with heterogeneous distribution (not random) of alleles and genotypes in space and time (Hamrick, 1982). If allelic and genotypic frequencies do not change across successive

generations, there is total absence of migration, mutation, selection and genetic drift. Through this principle it is possible to calculate the frequency of genotypes from the frequencies of the alleles.

There are ranges of tools utilized to study diversity in plants; which mainly include morphological, biochemical and molecular tools. Each tool has its own advantages and disadvantages. However, molecular markers are highly preferable and reliable tools to study polymorphism (or genetic diversity) among and within populations.

5. MOLECULAR MARKERS

According Matioli and Passos-Bueno (2001), genetic variability is very important tool for researchers in their studies when they wish to verify affinities and boundaries between species, detect modes of reproduction and population structure, estimate migration and dispersion levels in populations and even to help identify the remains of endangered species.

Through the use of molecular markers, advanced techniques have been employed in research allowing researchers to achieve greater efficiency in genetic diversity studies. In the molecular analysis technologies, DNA variability determines reference points on chromosomes, technically termed as "molecular markers". The revolution in this plan began with the discovery and use of isozymes from the year 1959, vastly expanding the number of genetic markers and enabling application of the technique virtually in all plant species (Ferreira and Grattapaglia, 1998).

Molecular markers can usually be used in genetic improvement programs, both quantitative and qualitative characters, either basic or applied research. Molecular markers offer numerous advantages over conventional phenotype (morphological) based alternatives as they are stable and detectable in all tissues regardless of growth, differentiation, development, or defense status of the cell, which are not confounded by

the environment, pleiotropic and epistatic effects (Semagn *et al*., 2006). However, molecular marker techniques differ from each other with respect to important features such as genomic abundance, level of polymorphism, locus specificity, reproducibility, technical requirements and cost (Agarwal *et al*., 2008).

5.1 Microsatellite (SSR) markers

Microsatellites are the choice for most studies in evolution, ecology and conservation, due to their high levels of polymorphism, high reproducibility, co-dominant nature and genome and locus-specificity (Powell *et al*., 1996; Pinheiro *et al*., 2009). Microsatellites (Litt and Luty, 1989), also known as simple sequence repeats (SSRs; Tautz *et al*., 1986), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs; Mcdonald and Potts, 1997), are the smallest class of simple repetitive DNA sequences. They are 2–8 bp (Armour *et al*., 1999); 1–6 bp (Goldstein and Pollock, 1997) or 1–5 bp (Schlotterer, 1998) repeats. SSRs are born from regions in which variants of simple repetitive DNA sequence motifs are already over represented (Tautz *et al*., 1986). The repeated sequence is often simple, consisting of two, three or four nucleotides (di-, tri-, and tetranucleotide repeats, respectively). The total number of repeats usually ranges between 10 and 100; and these markers often present high levels of polymorphism, particularly when tandem repeats number is ten or greater (Queller *et al*., 1993). SSR allelic differences are, therefore, the results of variable numbers of repeat units within the microsatellite structure and provide highly polymorphic banding patterns, which are often generated by PCR amplification of unique loci using discriminatory primers sets (Farooq and Azam, 2002).

Microsatellite markers have become quite useful in various aspects of molecular genetic studies in the past decade (Amsellem *et al*., 2001; Ashley *et al*., 2003). SSRs are

desirable because they are often co-dominant, highly reproducible, frequent in most eukaryotes, and reveal high allelic diversity (Mohan *et al*., 1997; Agarwal *et al*., 2008).

In macaúba, limited number of works has been conducted in the past since it was recognized as potential species only in recent years. Although, a dominant marker (RAPD) was used to study genetic diversity in macaúba populations (Oliveira *et al*., 2012), after the development of microsatellite markers by Nucci *et al*. (2008), more number of studies have been initiated and being realized in macaúba such as population structure and genetic diversity. However, the number of polymorphic SSR markers developed and characterized for macaúba is limited. Only eight SSRs have been characterized (Nucci *et al*., 2008) from an enriched genomic library containing a total of 77 sets of microsatellites (Nucci, 2007). In the characterization, the results demonstrated that the markers are promising in future population genetic studies to obtain information that will help to understand the evolutionary dynamics of the species and design conservation and breeding strategies. Nucci *et al*. (2008) reported genetic drift in most of the populations studied as a result of factors such as clearing of natural forests due to extractivism, climate change and land use policy, which lead to a recommendation of *ex situ* conservation ensuring use of its genetic resources for future breeding programs.

5.2 Development of microsatellite markers and transferability among species

A detail review by Semagn *et al*. (2006) discusses how SSR markers are developed and used in several studies. SSR primers are developed by cloning random segments of DNA from the target species. These are inserted into a cloning vector, which is in turn, implanted into *Escherichia coli* bacteria for replication. Colonies are then developed, and screened with single or mixed simple sequence oligonucleotide probes that will hybridize to a microsatellite repeat, if present on the DNA segment. If

positive clones for microsatellite are obtained from this procedure, the DNA is sequenced and PCR primers are chosen from sequences flanking such regions to determine a specific locus. This process involves significant trial and error on the part of researchers, as microsatellite repeat sequences must be predicted and primers that are randomly isolated may not display polymorphism (Queller *et al*., 1993; Jarne and Lagoda, 1996).

The next step is to select the best candidate markers and then to optimize conditions for their amplification. Optimization of microsatellite systems involves a more or less comprehensive survey of PCR conditions for amplification of candidate loci. This is done to adequately balance the often conflicting requirements for high specificity and high intensity of amplification products of microsatellite loci which are more common in some organisms than in others and screening may produce few useful loci in some species (Cooper, 1995). The efficiency of microsatellite marker development depends on the abundance of repeats in the target species and the ease with which these repeats can be developed into informative markers. When researchers are isolating plant microsatellites, about 30% of the sequenced clones, on average, can be lost due to the absence of unique microsatellites. Of those sequences that contain unique microsatellites, a number of the clones in a library can contain identical sequences (and hence there is a level of redundancy) and/or chimeric sequences (i.e., one of the flanking regions matches that of another clone). Hence, at each stage of SSR development, there is the potential to 'lose' loci, and hence the number of loci that will finally constitute the working primer set will be a fraction of the original number of clones sequenced (Squirrell *et al*., 2003). The conversion of microsatellite-containing sequences into useful markers can be quite difficult, especially in species with large genomes (Pfeiffer *et al*., 1997; Song *et al*., 2002). The low conversion rates of primer pairs to useful markers in these species are due to the high level of repetitive DNA

sequences in their genomes. The recovery rate for useful SSR primers is generally low due to different reasons: such as the primer may not amplify any PCR product; the primer may produce very complex, weak or nonspecific amplification patterns; and the amplification product may not be polymorphic.

SSRs are now the marker of choice in most areas of molecular genetics as they are highly polymorphic even between closely related lines, require low amount of DNA, can be easily automated for high throughput screening, can be exchanged between laboratories, and are highly transferable between populations (Gupta *et al*., 1999). The major constraint of using SSR markers from genomic libraries is the high development cost and effort required to obtain working primers for a given study species. This has restricted their use to only a few of the agriculturally important crops. A new alternative source of SSRs development from expressed sequence tag (EST) databases has been utilized (Kota *et al*., 2001; Kantety *et al*., 2002; Michalek *et al*., 2002). With the availability of large numbers of ESTs and other DNA sequence data, development of EST-based SSR markers through data mining has become a fast, efficient, and relatively inexpensive compared with the development of genomic SSRs (Gupta *et al*., 2003). This is due to the fact that the time-consuming and expensive processes of generating genomic libraries and sequencing of large numbers of clones for finding the SSR containing DNA regions are not needed in this approach (Eujayl *et al*., 2004). However, the development of EST-SSRs is limited to species for which this type of database exists. Furthermore, the EST-SSR markers have been reported to have lower rate of polymorphism compared to the SSR markers derived from genomic libraries (Cho *et al*., 2000; Scott *et al*., 2000; Eujayl *et al*., 2002; Chabane *et al*., 2005).

Hence, the wide application of SSRs has been limited in some species by the fact that the development of these markers is laborious and costly (Steinkellner *et al*., 1997; Barbará *et al*., 2007). According to Rossetto (2001), for new species in which

establishment cost of genomic library is high and EST is limited, cross-species amplification or transferability of SSR markers can be used as an alternative means for short term solutions. Closely related species are more likely to share SSR priming sites than more distantly related ones. Close to 90% of species/ primer combinations tested within subgenera were successful; a much higher rate than between genera attempts (just over 35%). It was reported also that when priming sites did transfer across species, the rate of polymorphism was still relatively high, even between evolutionary distant species (Rossetto, 2001).

Research on species relationships has increasingly focused on assessing the ability of SSR primers to amplify the same loci across different species and genera (Smulders *et al*., 1997; Steinkellner *et al*., 1997; Witsenboer *et al*., 1997). Besides, molecular ecologists increasingly require universal markers that can be readily transferred between species which facilitate comparisons among closely related taxa for addressing the mechanisms involved in population divergence and speciation (Barbará *et al*., 2007). However, there is a challenge that the distribution of cross-species transferability is highly uneven across taxa, being greater in animals and highly variable in flowering plants (Barbará *et al*., 2007).

Cross-species amplification of any DNA sequence is inversely related to the evolutionary distance between two species (Steinkellner *et al*., 1997). The successful rate of cross transferability for SSR primers within plants have been reported to average 89.8% at sub-genus; 76.4% at genus, and 35.2% at family level (Rossetto, 2001). However, the capacity of SSR primers to amplify microsatellite loci in different species is an advantage to facilitate comparative studies across related species, thus making the results directly comparable and the research more cost effective (Roa *et al*., 2000; Pinheiro *et al*., 2009).

7. REFERENCES

Agarwal M, Shivastava N and Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports* 27: 617-631.

Aguiar FFA (1988) Caracterização morfológica das principais espécies de palmeiras exóticas na cidade de São paulo. Dissertação Mesterado, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

Amsellem LC, Dutech and Billotte N (2001) Isolation and characterization of polymorphic microsatellite loci in *Rubus alceifolius* Poir (Rosaceae), an invasive weed in La Reunion island. *Molecular Ecology Notes* 1: 33-35.

Armour JAL, Alegre SA, Miles S, Williams LJ and Badge RM (1999) Minisatellites and mutation processes in tandemly repetitive DNA. In: Goldstein DB, Schlotterer C (ed.) *Microsatellites: evolution and applications.* Oxford, pp. 24–33.

Ashley MV, Wilk JA, Styan SMN, Craft JK, et al. (2003) High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill (Rosaceae). *Theoretical and Appllied Genetics* 107: 1201-1207.

Averdunk K, Zelt T, Golka P, Höpfner M, Müller C and Bettermann I (2013) Macauba –Sustainable Palm Oil, Results of the Feasibility Study of the Leuphana University of Lüneburg- *Executive Summary*. Available at http://www.leuphana.de.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF and Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology* 16: 3759–3767.

Cesar E (2003) Biodiesel no tanque. Pesquisa Fapesp, 94 ed., São Paulo.

Chabane K, Ablett GA, Cordeiro GM, Valkoun J and Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resource and Crop Evolution* 52:903–909.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Parl WD, Ayers N and Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 100:713–722.

Chuba CAM, Tommaselli MAG, Santos WL and Sanjinez-Argandoña EJ (2008) Parâmetros biométricos dos cachos e frutos da bocaiuva. In: *Congresso Brasileiro de Fruticultura*, Vitória.

Cooper G (1995) Analysis of genetic variation and sperm competition in Dragon flies. D.Phil.Thesis, Oxford University.

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC an Mian MAR (2004) Medicago truncatula EST-SSRs reveal crossspecies genetic markers for Medicago spp. *Theoretical and Applied Genetics* 108:414–422.

Eujayl I, Sorrells ME, Wolters P, Baum M and Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics* 104:399–407.

Farooq S and Zam F (2002) Molecular Markers in Plant Breeding-I: Concepts and Characterization. *Pakistan Journal of Biological Sciences* 5: 1135-1140.

Ferreira ME and Grattapaglia D (1998) Introducción al Uso de Marcadores moleculares en el Análisis Genético. Brasilia: EMBRAPA-CENARGEN, p. 220..

Goldstein DB and Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Joural of Hered* 88: 335–342.

Grin−Germplasm Resources Information Network. Available at http://www.arsgrin.gov/cgi on 12/08/2013.

Govindaraj M, Vetriventhan, M and Srinivasan (2015) Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives. *Genetics Research International* Doi: http://dx.doi.org/10.1155/2015/431487.

Gupta PK, Rustgi S, Sharma S, Sing R, Kumar N and Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *MolecularGenetics and Genomics* 270: 315–323.

Gupta PK, Varshney RK, Sharma PC and Ramesh B (1999) Molecular markers and their applications in wheat breeding. *Plant Breeding* 118:369–390.

Henderson A, Galeano G and Bernal R (1995) Field guide to palms of the Americas, New Jersey: Princepton University, pp. 166-167.

Hamrick JL (1982) Plant population genetics and evolution. *American Journal of Botany* 69 1685-1693.

Holanda A (2004) Biodiesel e a Inclusão Social. Brasília: Câmara dos Deputados, Coordenação de Publicações, Brasilia, p. 200.

Jarne P and Lagoda P (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* 11: 424–429.

Kantety RV, Rota ML, Matthews DE and Sorrells ME (2002) Data mining for simple-sequence repeats in expressed sequence tags from barley, maize, rice, sorghum, and wheat. *Plant Molecular Biology* 48:501–510.

Kota R, Varshney RK, Thiel T, Dehmer KJ and Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* 135:145–151.

Lima ES, Felfili JM, Marimon BS and Scariot A (2003) Diversidade, estrutura e distribuição espacial de palmeiras em um cerrado *sensu stricto* no Brasil Central. *Revista Brasileira de Botânica* 26: 361-370.

Litt M and Luty JM (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44: 397–401.

Lorenzi H (1992) Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Nova Odessa, São Paulo, Plantarum.

Lorenzi H (1996) Palmeiras do Brasil: exóticas e nativas. Nova Odessa: Editora Plantarum, pp. 1-20.

Lorenzi H (2004) Palmeiras Brasileiras e exóticas cultivadas, Instituto Plantarum, p. 1-20.

Matiolo SR and Passos-Bueno MRS (2001) Métodos baseados em PCR para análise de polimorfismos de ácidos nucléicos. In: Matioli SR.(ed). *Biologia Molecular e Evolução*. Ribeirão Preto: Holos, pp. 153-161.

Mcdonald DB, Potts WK (1997) DNA microsatellites as genetic markers for several scales. In: Mindell DP (ed.) *Avian molecular evolution and systematics*. San Diego, pp. 29–49.

Michalek V, Weschke W, Pleisner KP and Graner A (2002) EST analysis in barley defines a unigene set comprising 4000 genes. *Theoretical and Applied Genetics* 104: 97–103.

Morcote-Rios G and Bernal R (2001) Remains of palms (Palmae) at archaeological sites in the New Word: a review. *The Botanical Review*, 67: 309-350.

Nei M (1978) Estimation of average heterozygosity and genetic distances from a small number of individuals. *Genetics* 89: 583–590.

New World Fruits −Database. Available at http://www.ipgri.cgiar.org on 06/07/2014.

Novaes RF (1952) Contribuição para o estudo do coco macaúba. Tese de Doutorado, Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo.

Nucci SM, Azevedo-Filho A, Colomo AC, Priolli GHR, Coelho MR, Mata TL and Zucchi IM (2008) Development and characterization of microsatellites markers from the macaw. *Molecular Ecology Resources* 8: 224–226.

Nucci SM (2007) Development, characterization and analyses of the use of microsatellite markers in macaw palm population genetics. MSc. Thesis, University of Sâo Paulo.

Oliveira DA, Júnior-Melo AF, Brandão MM, Rodrigues LA, Menezes EV and Ferreira PRB (2012) Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais. *Genetics and Molecular Research* 11: 531–538.

Oliveira FAM (2006) A produção de óleos vegetais de macaúba e seus co-produtos na região metropolitana de Belo Horizonte. In: *3º Congresso Brasileiro de Plantas Oleaginosas, óleos, Gorduras e Biodiesel*, Varginha.

Pfeiffer A, Olivieri AM, Morgante M (1997) Identification and characterization of microsatellites in Norway spruce (*Picea abies* K.). *Genome* 40:411–419.

Pinheiro F, Palma-Silva C, Barros F and Cozzolino S (2009) Cross-amplification and characterization of microsatellite loci for the Neotropical orchid genus *Epidendrum*. *Genetics and Molecular Biology* 32: 337–339.

Pinto LR (2001) Análise da estrutura genética das populações de milho (*Zea mays* L.) BR-105 e BR-106 e respectivos sintéticos IG-3 e IG-4 por meio de microssátelites. Tese Doutorado, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Moleculur Breeding* 2: 225–238.

Queller DC, Strassman JE and Hughes CR (1993) Microsatellites and Kinship. *Trends in Ecology & Evolution* 8: 285-288.

Ramos MIL, Ramos Filho MM, Hiane PA, Braga Neto, JA, Siqueira EMA (2008) Qualidade nutricional da polpa de bocaiúva *Acrocomia aculeata*(Jacq.) Lodd. *Ciência e Tecnologia de alimentos* 28: 90-94.

Ratter JA, Bridgewater S, Atkinson R and Ribeiro JF (1996) Analysis of the floristic composition of the Brazilian cerrado vegetation. *Edinburgh Journal of Botany* 53: 153-180.

Read RW (1966) New chromosome counts in the Palmae. *Principles* 10: 55-61.

Roa AC, Chavarriaga-Aquire P, Dugue MC, Maya MM, Bonierbale MW, Iglesias C and Tohme JM (2000) Cross species-amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationships. *American Journal of Botany* 87: 1647–1655.

Rosetto M (2001) Sourcing of SSR markers from related plant species. In: Henry RG (ed.) *Plant Genotyping: The DNA finger printing of plants*. New York. CAB International, pp. 211–224.

Scariot AO, Lleras E and Hay JD (1991) Reproductive biology of the palm *Acrocomia aculeata* in Central Brazil. *Biotropica* 23: 12–22.

Scariot AO, Lleras E and Hay JD (1995) Flowering and fruiting phenologies of the palm *Acrocomia aculeata*: patterns and consequences. *Biotropica* 27: 168-173.
Schlotterer C (1998) Microsatellites. In: Hoelzel AR (ed) *Molecular genetic analysis of populations: A practical approach.* Oxford, pp. 237–261.

Semagn K, Bjørnstad Å and Ndjiondjop MN (2006) A review on an overview of molecular marker methods for plants. *African Journal of Biotechnology* 5: 2540-2568.

Sharma AK, Sarkar SK (1956) Cytology of different species of palms and its bearing on the solution of the problems of phylogeny and speciation. *Genetica* 28: 361–488.

Silva JA, Silva DB, Junqueira NTV and Andrade LRN (2001) Frutas nativas dos cerrados. *Nrasília*: EMBRAPA-CPAC. p. 166.

Silva JC (1994) Macaúba: fonte de matéria-prima para os setores alimentício, energético e industrial. Trabalho de conclusão da disciplina Cultivo de essências exóticas e nativas. Departamento de Engenharia Florestal, Universidade Federal de Viçosa.

Smulders MJM, Bredemeijer G, Rus-Kortekass W, Arens P and Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. *Theoretical and Applied Genetics* 97: 264–272.

Song QJ, Fickus EW and Cregan P (2002) Characterization of trinucleotide SSR motifs in wheat. *Theoretical and Applied Genetics* 104:286–293.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M and Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* 12:1339–1348.

Steinkellner H, Lexer C,Turetschek E and Glossl J (1997) Conservation of (GA)n microsatellite between *Quercus* species. *Molecular Ecology* 6:1189–1194.

Tautz D, Trick M and Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652−656.

Weir BS (1996) Genetics data analysis II: methods for discrete population genetic data. Suderland, MA: Sinauer Associates, p. 455.

CHAPTER 1

[1]CROSS-SPECIES AMPLIFICATION AND CHARACTERIZATION OF NEW MICROSATELLITE MARKERS FOR THE MACAW PALM, *Acrocomia aculeata* (Arecaceae).

ABSTRACT

Microsatellites (Simple Sequence Repeats−SSRs) are useful molecular markers allowing for efficient conservation and sustainable use of genetic resources of plant species. Development of SSR marker system for new species is a very expensive task and time consuming. Cross-species amplification of microsatellite loci is considered as a cost-effective approach for developing microsatellite markers for new species. The aim of this work was to examine the transferability of some SSR markers of two Arecaceae species (*Astrocaryum aculeatum* and *Elaeis oleifera*), in *Acrocomia aculeata*. A total of 34 SSR markers (14 from *A. aculeatum* and 20 from *E. oleifera*) were evaluated for the trsnaferability using 192 *A. aculeata* individuals. Fifteen (44%) of the markers successfully amplified the genomic DNA in *A. aculeata*, of which 4 (26%) were polymorphic detecting a range of 3 to 8 alleles with an average of 4.5 per locus. High average percentage of polymorphic loci (P = 71.2%) per provenance was obtained within a range of 57–100% detecting genetic variation in *A. aculeata* germplasm collections. The polymorphic markers detected a positive inbreeding coefficient (F > 0) per locus revealing heterozygote deficiency in the accessions that were analysed. As the cross-amplification was at family level, in which the taxonomic distance is relatively wider between the sources (*A. aculeatum* and *E. oleifera*) and the target (*A. aculeata*) species, the amplification success was relatively low. However, the results are promising and implicated that high cross-amplification success could be achieved at species or genus level in *A. aculeata*. The markers will contribute towards the domestication of the potential macaw palm through realizing various studies such as population genetics, germplasm characterization, genetic improvement and conservation.

Key words: biodiesel; domestication; macauba; null alleles; SSRs; transferability

[1] Published online in Journal of Plant Genetic Resource (PGR)- Characterization and utilization (doi: http://dx.doi.org/10.1017/S1479262115000179)

AMPLIFICAÇÃO CRUZADA E CARACTERIZAÇÃO DE NOVOS MARCADORES MICROSSATÉLITES PARA A MACAÚBA, *Acrocomia aculeata* (Arecaceae).

RESUMO

Os microssatélites (Simple Sequence Repeats−SSR) são marcadores moleculares úteis que auxiliam na conservação eficiente e uso sustentável dos recursos genéticos das espécies de plantas. No entanto o desenvolvimento de marcadores SSR para uma nova espécie é uma tarefa demorada e muito dispendiosa. A amplificação cruzada de locos SSR é considerada uma alternativa para o desenvolvimento de desses marcadores para novas espécies. O objetivo deste trabalho foi avaliar a possibilidade de transferência de alguns marcadores SSR de duas espécies de Arecaceae (*Astrocaryum aculeatum* e *Elaies oleifera*), em *Acrocomia aculeata*. Um total de 34 marcadores SSR (14 de *A. aculeatum* e 20 de *E. oleifera*) foram avaliados na transferência usando 192 acessos de *A. aculeata*. Do total de marcadores analisados, 15 (44%) amplificaram com sucesso o DNA genómico de *A. aculeata*, dos quais 4 (26%) foram polimórficos e produziram de 3 a 8 alelos, com uma média de 4,5 por loco. Alta média de percentagem de locos polimórficos (P = 71,2%) por procediência foi obtido dentro de um intervalo de 57 a 100%, detectando uma variação genética na coleção de germoplasma de *A. aculeata*. Os marcadores polimórficos detectaram um coeficiente de endogamia positivo (F > 0) por loco, mostrando deficiência de heterozigose nos acessos que foram analisados. Como a amplificação cruzada foi em nível da família, em que a distância taxonômica é relativamente elevada entre as espécies-fonte (*A. aculeatum* e *E. oleifera*) e a espécie-alvo (*A. aculeata*), o sucesso de amplificação foi relativamente baixa. No entanto, os resultados são promissores e implica que uma alta taxa de sucesso em amplificação cruzada poderia ser alcançada em nível de gênero ou espécie de *Acrocomia*. Os marcadores apresentados contribuirão para a domesticação da macaúba através da viabilização de vários estudos, tais como genética de populações, caracterização de germoplasma, melhoramento genético e conservação.

Palavras-chave: alelos nulos; biodiesel; domesticação; macaúba; SSRS; transferibilidade

1. INTRODUCTION

Microsatellites (Simple Sequence Repeats, SSRs) are the choice of many studies because of their high levels of polymorphism, reproducibility, co-dominant nature and genome and locus-specificity (Powell *et al*., 1996; Pinheiro *et al*., 2009).

The high transfer rate of SSRs between related species is an advantage to save a large portion of development time and costs (Barbará *et al*., 2007). Hence, sourcing of SSR primers from related species could be a cost-effective alternative to develop primers especially for new species where abundant sequence data is not available (Kantety *et al*., 2002) and resources for developing new SSR primers are limited (Roa *et al*., 2000; Pinheiro *et al*., 2009).

The Macaw palm, *A. aculeata* (Jacq.) (Lodd. ex Mart.)−Arecaceae (2n=2x=30) (Abreu *et al*., 2011), is an emerging perennial palm native to South America. It is monoecious and self-compatible with androgynous inflorescence, which bears a mixed reproductive system, with a predominance of out-crossing (Scariot *et al*., 1991; Abreu *et al*., 2012). The combination of two pollination strategies (entomophily and anemophily) with flexible reproductive systems (cross- and self-pollination) suggests that *A. aculeata* can be highly successful in the colonization of new areas, as evidenced by its ample distribution in Brazilian biomes and in the rest of the Neotropics (Scariot *et al.*, 1991).

The species is known for its high oil-producing potential, thereby becoming sources of biofuel for both aviation and automotive sectors (Lanes *et al*., 2014). It produces up to 25 tonnes/ha of fruits, which can be processed to 4000 kg of vegetable oil and the resulting solid waste can be transformed into charcoal and nutritious cakes to generate energy and feed livestock (Tickel, 2000; Moura *et al*., 2009). The biochemical properties of the oil are suitable for cosmetic industries and for biodiesel production (Fortes and Baugh, 1999; Bora and Rocha, 2004; Hiane *et al*., 2005). Moreover, this palm has environmental benefits as it can foster in impoverished soils and drought-

prevailing areas (Motoike and Kuki, 2009). Hence, *A. aculeata* is a suitable option for production of biodiesel among the common food-based oleaginous plants such as soybean, sunflower and oil palms (Teixeira, 2005).

However, *A. aculeata* has risks of predatory extractivism in natural populations, unsustainable land use and climate change which potentially threaten its natural genetic diversity (Faleiro *et al*., 2008; Ribeiro *et al*., 2011). Our study focused on the development of new microsatellite markers for *A. aculeata* sourcing SSRs from related palm species to contribute for the conservation of genetic resources of this noble palm.

Therefore, this study was carried out to evaluate the transferability of some SSR markers from *Astrocaryum aculeatum* and *Elaeis oleifera* to *A. aculeata*. It was also performed to identify and characterize some polymorphic SSR markers from sets of markers previously developed for *A. aculeata*.

2. MATERIALS AND METHODS

2.1 Plant material and DNA isolation

Leaf samples of 192 *A. aculeata* germplasm accessions were obtained from the *ex situ* plant collection, Macaúba Active Germplasm Bank (BAG – Macaúba respository #: 084/2013/CGEN/MMA) located in the experimental farm of the Universidade Federal de Viçosa in the municipality of Araponga ($20^{o}40'1''S$, $42^{o}31'15''W$), State of Minas Gerais, Brazil (Figure 1). The accessions, obtained from six States in Brazil were developed from seeds germinated using a pre-germination protocol as described in patent INPI 014070005335 (Motoike *et al*., 2007). The accessions represent six provenances containing 41 different families in which each family has a maximum of five individuals (Table 1).

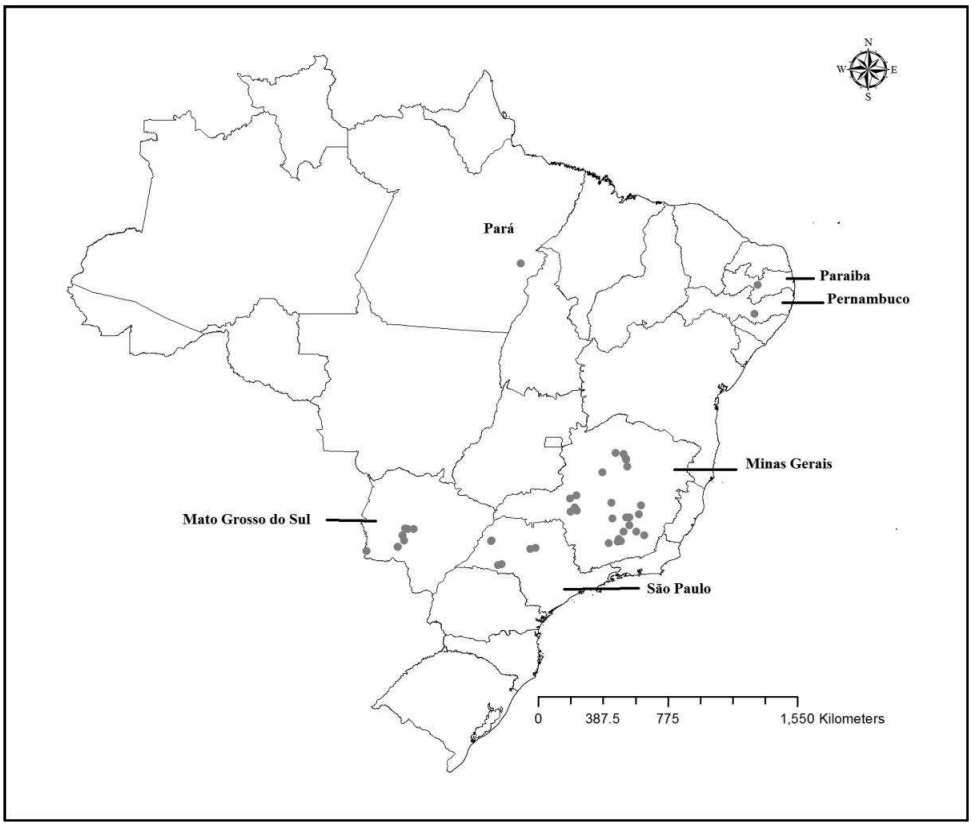

Figure 1. Map of Brazil showing the six geographical States, where the original plant materials were collected. The States include Pará; Pernambuco; Paraiba; São Paulo; Minas Gerais and Mato Grosso do Sul. Araponga is a city in Minas Gerais State, where the germplasm bank is located in which the experimental plant materials were obtained.

Genomic DNA (gDNA) was isolated from each individual sample according to CTAB (Cetyl Tri-methyl Ammonium Bromide) method (Doyle and Doyle, 1990) with some modifications as described by Lanes *et al*. (2013). Isolated DNA samples were quantified with Multiscan^{TM} GO Microplate Spectrophotometer (Thermo Fisher Scientific OY, Ratasite, Finland) at absorbance of 260 and 280 nm. The integrity of the DNA samples was confirmed by 2% agaros gel and the working concentration was adjusted to 30 ng/µl.

Table 1. List of 41 *Acrocomia aculeata* families assessed in the macaúba active germplasm bank (BAG-Macauba)

No.	Family code	State/ Provenance[a]	Coordinates[b]		No.	Family code	State/ Provenance[a]	Coordinates[b]	
			Latitude	Longitude				Latitude	Longitude
1	BGP99	PA	S 06 03 58.0	W 49 33 39.0	22	BGP11	MG	S 19 14 01.2	W 43 03 28.4
2	BGP82	PE	S 07 14 23.0	W 36 46 55.0	23	BGP9	MG	S 19 33 12.0	W 46 51 10.1
3	BGP124	PB	S 08 48 49.0	W 36 57 14.0	24	BGP78	MG	S 18 51 25.6	W 46 52 55.2
4	BGP51	SP	S 21 32 04.6	W 48 44 24.7	25	BGP37	MG	S 18 40 51.3	W 46 33 41.4
5	BGP34	SP	S 22 25 10.8	W 50 34 43.1	26	BGP33	MG	S 19 19 40.3	W 46 38 11.5
6	BGP47	SP	S 22 29 14.2	W 50 46 16.2	27	BGP21	MG	S 19 31 15.9	W 46 31 42.2
7	BGP20	MG	S 16 39 52.7	W 43 53 58.9	28	BGP2	MG	S 20 39 20.4	W 43 18 45.2
8	BGP27	MG	S 16 21 20.7	W 44 25 30.5	29	BGP76	MG	S 19 41 51.4	W 43 11 27.7
9	BGP22	MG	S 17 25 54.0	W 45 08 59.5	30	BGP25	MG	S 17 06 54.6	W 43 49 16.4
10	BGP16	MG	S 16 26 07.6	W 44 00 50.5	31	BGP64	MG	S 16 44 12.7	W 43 51 54.9
11	BGP49	MG	S 20 38 58.0	W 44 01 15.5	32	BGP105	MS	S 20 29 52.5	W 55 18 39.3
12	BGP10	MG	S 21 03 12.9	W 44 16 28.2	33	BGP102	MS	S 20 30 38.6	W 55 37 59.7
13	BGP68	MG	S 21 11 27.6	W 44 19 29.7	34	BGP104	MS	S 20 27 55.9	W 55 46 41.7
14	BGP3	MG	S 21 09 52.2	W 44 08 49.5	35	BGP117	MS	S 20 27 56.5	W 55 46 38.2
15	BGP51	MG	S 21 17 20.5	W 44 49 12.6	36	BGP118	MS	S 20 50 22.3	W 55 54 53.3
16	BGP5	MG	S 19 05 02.0	W 44 39 13.9	37	BGP112	MS	S 20 50 16.5	W 55 54 51.8
17	BGP14	MG	S 19 56 29.0	W 44 36 12.0	38	BGP106	MS	S 21 28 42.3	W 56 10 03.6
18	BGP18	MG	S 19 52 34.0	W 43 52 20.5	39	BGP92	MS	S 21 28 45.7	W 56 10 06.6
19	BGP24	MG	S 19 53 20.2	W 43 41 11.5	40	BGP103	MS	S 21 42 04.8	W 57 50 39.0
20	BGP1	MG	S 20 17 42.6	W 43 42 30.9	41	BGP119	MS	S 21 42 06.0	W 57 50 32.4
21	BGP52	MG	S 20 50 13.1	W 42 54 27.3					

[a]States include: PA=Pará, PE=Pernambuco, PB=Paraiba, SP=São Paulo, MG=Minas Gerais and MS=Mato Grosso do Sul.
[b]Coordinates are in degrees, minutes and seconds for both the latitude (S=South) and longitude (W=West).

2.2 Condition of Polymerase Chain Reaction (PCR) and electrophoresis

The PCR was performed in a 20 µL total reaction volume containing 30 ng of *A. aculeata* genomic DNA; 1xPCR buffer; 250 µM of each dNTPs; 0.15 µM of forward and reverse primer; 3 mM of $MgCl_2$; and 1 U of *Taq* DNA polymerase (Nucci et al., 2008). Some modifications were made by increasing the concentration of $MgCl_2$ to 4 mM for markers Aac04 and Aac12 for better amplification (Table 2). Totally, 20 SSR markers, which were originally developed for *E. oleifera* (Zaki et al., 2012), 14 SSR markers developed for *A. aculeatum* (Ramos et al., 2012) and three identified from *A. auleata* (Nucci, 2007) were used in the reactions (Table 2).

The amplification cycles were carried out on a PCR thermal cycler (Applied Biosystem®Verti®cycler, Thermo Fisher Scientific Brand, USA) programmed as follows: initial denaturation step at 94^{o}C for 5 min followed by 30 cycles of denaturation at 94^{o}C for 1 min; annealing at primer-specific annealing temperature for 1 min (Table 2); extension at 72^{o}C for 1 min and the final extension at 72^{o}C for 8 min (Nucci et al., 2008). PCR products were denatured in a bromophenol blue dye solution at 95^{o}C for 5 min in the thermal cycler immediately before electrophoresis on 6% polyacrylamide gel in 1XTBE (Tris-Borate-EDTA) buffer solution at 60W for 1 hr and 40 min.

2.3 Polyacrylamide gel staining

After electrophoresis, the PCR products were visualized in polyacrylamide gels stained with silver nitrate ($AgNO_3$) according to Brito *et al*. (2010). The gels were immersed and agitated in several coloring steps in different solutions at different concentrations and durations until all allelic bands were totally visible for evaluation. Finally, the stained gels were allowed to dry in the air and scanned for documentation and DNA fragments were scored as co-dominant alleles for data analysis.

Table 2. List of SSR primer sequences used to amplify the target microsatellite loci in the cross amplification.

Locus	Forward primer (5´-3´)	Reverse primer (5´-3´)	Repeat motif	Accession name	T_a (°C)	Source[a]
Aac04	F: GCATTGTCATCTGCAACCAC	R: GCAGGGGCCATAAGTCATAA	(GT)7(GA)16	GF111930	60	1
Aac08	F: CGCACGTACACACACACACAT	R: GCAGGGGCCATAAGTCATAA	(CA)11	GF111934	57	1
Aac10	F: AGCCGTGAGTGAACTGCTTT	R: AAGCCCAAACTTCTTCCTCG	(CT)7	GF111936	60	1
Aac11	F: AAAGGAACAACCCAAGAGGG	R: TGGGGAGTGGACGTAAGTGT	(AC)5	GF111937	60	1
Aac12	F: GCTCTGTAATCTCGGCTTCCT	R: TCCAGTTCAAGCTCTCTCAGC	(GC)5(AC)3	GF111938	60	1
Aac13	F: CTAGACAACCCAAGAGAGGGG	R: TTGGAGAGTGGATGTAGGTGC	(CA)7	GF111939	60	1
sMo00020	F: CCTTTCTCTCCCTCTCCTTTTG	R: CCTCCCTCCCTCTCACCATA	(AG)15	Pr009947964	58	2
sMo00027	F: TTACAGTTGAGGCAGTATGTCAAT	R: CTGTATGTCAAACCTTCTGCAC	(TC)14	Pr009947965	50	2
sMo00055	F: GGCATTTCAGATAACGACAAA	R: GCACCCAAGTCTCTCTACCTC	(GA)11	Pr010315684	54	2
sMo00130	F: TAAGCAAAAGATCAGGGCACTC	R: GCTGGTGAAAATAGGTTTACAAAG	(AAG)11	Pr009947968	56	2
sMo00137	F: AGGAAGGAGAAGGAGATGAACAG	R: CTTTGGATTTGAGCAGAGGAAG	(AAAT)6	Pr010315687	54	2
sMo00018	F: TTAAATGAGAGAGAGACGAGGAC	R: TGGAGCCATGAGAAAGAGTA	(CT)14	Pr009947963	54	2
sMo00141	F: ACTTGACATACAGGTTCCACTGA	R: CCTGCTACCTCCTAATTCTATCAAA	(TTCTT)5	Pr010317029	56	2
sMo00147	F: TACCCAATCCCACCGAGTTA	R: CGTCTCCACTGAACCACAAAA	(AAAAG)5	Pr010317030	54	2
sMo00161	F: ACTGTTTCGTCAAGCATTTG	R: ATCAAGAGAAGGTCGTGTCAG	(TG)8(AG)8	Pr010317032	54	2
Aacu38	F: TTCTCAGTTTCGTGCGTGAG	R: GGGAGGCATGAGGAATACAA	(TC)15	–	56	3
Aacu45	F: CAGACTACCAGGCTTCCAGC	R: TCATCATCGCAGCTTGACTC	(CGAC)5	–	56	3
Aacu74	F: TACTGTTGTGCCAAGTCCCA	R: GAGCACAAGGGGGATATCAA	(CA)15	–	56	3

[a] 1, Ramos *et al*. (2012); 2, Zaki *et al*. (2012); 3, Nucci (2007); T_a= annealing temperature

2.4 Data analyses

Co-dominant data were analyzed using the GENES statistical software program (Cruz, 2013) to estimate allelic diversity, heterozygosity and polymorphism level of the SSR markers. The number of alleles per locus (*A*) was determined by quantifying the number of different alleles amplified by each marker analyzed with 192 individuals. The total number of alleles per provenance (N_t) was determined by summing the number of alleles amplified by each locus (*A*). Hence, the average number of alleles per provenance (N_a) was calculated from the total number of alleles detected in the provenances by the loci that were analyzed (Cruz *et al*., 2011). Effective numbers of alleles (N_e), which are used to make sampling in successive generations of a given population, were determined by quantifying the number of alleles amplified by the polymorphic loci out of the total number of loci analyzed with a criterion that alleles have a frequency of < 0.95 (Cole, 2003).

The heterozygosity level of each locus was determined. Expected heterozygosity (H_E) per locus was estimated from the frequency of alleles detected per locus; while observed heterozygosity (H_o) was determined by calculating the number of heterozygotes observed out of the total number of individuals analyzed (Nei, 1978). H_E and H_o were used to estimate the inbreeding coefficient (F) per locus to determine whether there are excess (or any deficiency) of heterozygotes per locus (Hartl and Clark, 1997).

The informativeness of the SSR markers was estimated in the analyses by calculating the polymorphic information content (PIC) of each locus, which was computed from the frequency of alleles detected per locus according to Bostein *et al*. (1980). To estimate the proportions of polymorphic loci detected in each provenance, the percentage of polymorphic loci (*P*) per provenance was determined by quantifying the number of polymorphic loci obtained out of the total number of loci analyzed based

on a criterion when a locus with most common alleles has a frequency of less than 0.95 (Cole, 2003).

The null allele test was performed to check for evidence of null alleles detected by the loci using the FreeNA (Chapuis and Estoup, 2007) computer program based on the Expectation Maximization (EM) algorithm as described in Dempster *et al*. (1977), which estimated the null allele frequency (*p)* for each locus across the provenances analyzed. A test for deviations from Hardy-Weinberg Equilibrium (HWE) was performed for each locus across the provenances at a significance level of α=0.05 in GENEPOP (Raymond and Rousset, 1995).

3. RESULTS

3.1 Cross-Amplification and Polymorphism

In the cross-amplification, 15 of the total SSR markers (44%) from *A. aculeatum* and *E. oleifera* successfully produced amplicons, and were able to amplify the gDNA in *A. aculeata* (Table 3). However, only four (26%) of the markers were polymorphic and they detected a range of three to eight alleles with average of 4.5 per locus (Table 3). Besides, three new polymorphic SSR markers were identified from sets of SSR markers previously designed for *A. aculeata*. Hence, a total of 38 alleles were amplified by the seven polymorphic SSR markers with average of 5.4 per locus (Table 3).

Alleles amplified by the polymorphic loci were detected in the six provenances at different proportions from 14 (36.8 %) to 32 (84.2 %) (Table 3). Out of the total number of alleles, a range of 2.4 to 6.8 effective numbers of alleles (N_e) with average of four per provenance was obtained (Table 3). Number of polymorphic loci across the provenances varied from 3.99 (*P*=57 %) to 7 (*P*=100 %) with average of 4.98 (*P*= 71.2 %) per provenance (Table 3). Observed heterozygosity (H_o) ranged from 0.01

(sMo00137) to 0.61 (Aac04) with average of 0.20 per locus, while expected heterozygosity (H_E) ranged from 0.09 (sMo00020) to 0.72 (Aac04) with average of 0.39 per locus (Table 3). A range of positive inbreeding coefficients (F) were obtained from 0.15 (Aac04) to 0.91 (sMo00137) averaging 0.57 per locus, while PIC of the markers varied from 0.08 (sMo00020) to 0.68 (Aac04) with average of 0.35 per locus (Table 3).

Table 3. Characteristics of 15 SSR markers identified from *Astrocarym aculeatum* (Ramos *et al*., 2012) and *Elaeis oleifera* (Zaki *et al*., 2012), which showed amplification in *Acrocomia aculeata* accessions and three SSR markers screened from sets of markers developed for *A. aculeata* (Nucci, 2007).

			Number of alleles amplified per provenance									
Locus	Amplicon (bp)	*A*	PA	PE	PB	SP	MG	MS	H_o	H_E	F	PIC
Aac04[a]	258-306	8	4	5	4	7	7	6	0.61	0.72	0.15	0.68
Aac08	428	1	1	1	1	1	1	1	-	-	-	-
Aac10	134	1	1	1	1	1	1	1	-	-	-	-
Aac11	302	1	1	1	1	1	1	1	-	-	-	-
Aac12[a]	229-247	4	2	1	2	3	2	3	0.06	0.31	0.81	0.27
Aac13	355	1	1	1	1	1	1	1	-			-
sMo00020[a]	242-250	3	1	1	1	3	2	3	0.03	0.09	0.67	0.08
sMo00027	293	1	1	1	1	1	1	1	-	-	-	-
sMo00055	273	1	1	1	1	1	1	1	-	-	-	-
sMo00130	276	1	1	1	1	1	1	1	-	-	-	-
sMo00137[a]	179-187	3	1	1	1	2	2	2	0.01	0.11	0.91	0.09
sMo00138	305	1	1	1	1	1	1	1	-	-	-	-
sMo00141	441	1	1	1	1	1	1	1	-	-	-	-
sMo00147	310	1	1	1	1	1	1	1	-	-	-	-
sMo00161	221	1	1	1	1	1	1	1	-	-	-	-
Aacu38[a]	316-346	6	4	2	2	4	6	6	0.13	0.64	0.80	0.58
Aacu45[a]	260-284	5	4	2	2	2	5	4	0.30	0.38	0.21	0.34
Aacu74[a]	278-313	9	1	2	2	6	8	6	0.26	0.45	0.42	0.42
N_t		38	17(44.7)	14(36.8)	14(36.8)	27(71.1)	32(84.2)	30(78.9)	-	-	-	-
N_a		5.4	2.4	2.0	2.0	3.9	4.7	4.3	-	-	-	-
N_e		-	3.5	2.8	2.4	4.2	6.8	4.3	-	-	-	-
P		-	3.99(57)	3.99(57)	4.97(71)	5.95(85)	3.99(57)	7(100)	-	-	-	-
Mean H_o, H_E, F and PIC		-	-	-	-	-	-	-	0.20	0.39	0.57	0.35

A, number of alleles per locus; H_o, observed heterozygosity; H_E, expected heterozygosity; F, inbreeding coefficient; PIC, polymorphic information content; N_t, total number of alleles over polymorphic loci and provenance (numbers in brackets are percentages); N_a, average number of alleles per polymorphic loci and per provenance; N_e, effective number of alleles per provenance; *P*, percentage of polymorphic loci in brackets (numbers outside brackets are numbers of polymorphic loci); Provenances: PA=Pará, PE=Pernambuco, PB=Paraiba, SP=São Paulo, MG=Minas Gerais, MS=Mato Grosso do Sul. [a] Rows refer to the polymorphic loci.

The polymorphism and heterozygosity levels of the seven polymorphic SSR markers that amplified the target microsatellite loci in *A. aculeata* were depicted in a figure (Figure 2). This figure demonstrated the allelic profiles of the loci amplified on 12 selected families of *A. aculeata* containing 48 accessions from the six provenances and elucidated the capability of the markers to distinguish between heterozygote and homozygote individuals in *A. aculeata*.

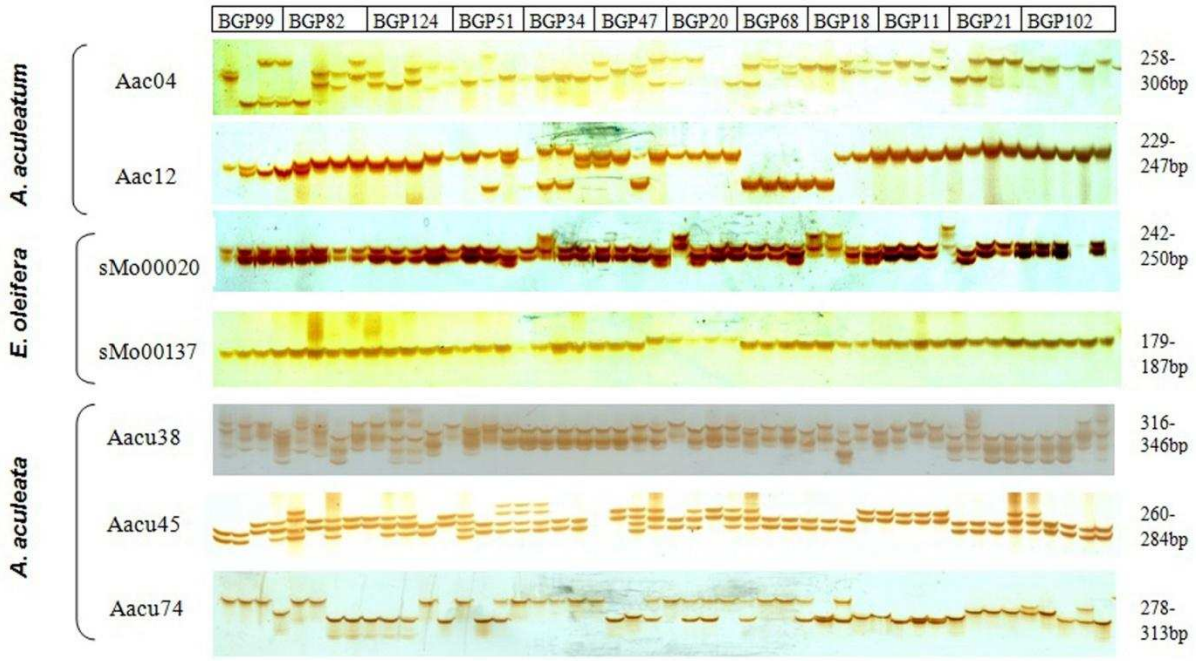

Figure 2. Allelic profiles showing cross-species amplifications of seven SSR loci on *Acrocomia aculeata* accessions. Loci include Aac04 and Aac12 (*Astrocarym aculeatum*); sMo00020 and sMo00137 (*Elaies oleifera*); and Aacu38, Aacu45 and Aacu74 (*Acrocomia aculeata*). The amplification is on 48 selected *A. aculeata* accessions of twelve families. BGP99 (Pará); BGP82 (Pernambuco); BGP124 (Paraiba); BGP51, BGP34 and BGP47 (São Paulo); BGP20, BGP68, BGP18, BGP11 and BGP21 (Minas Gerais) and BGP102 (Mato Grosso do Sul). Allele amplicon size (bp) indicated for each locus.

The test for evidence of null alleles identified the presence of null alleles across the loci presented in the form of allelic frequency (*p)* in which the intensity of the null alleles varied among the loci and the provenances analyzed (Table 4). A complementary test for HWE showed significant deviations of the loci in certain provenances suspected

of having null alleles (Table 4). The combination of the latter two test results showed that loci with strong null allele frequency are significantly deviated from HW (Table 4).

Table 4. Locus by provenance table of estimated null allele frequencies (p)[a] and Hardy-Weinberg deviation test.

Locus	PA		PE		PB		SP		MG		MS	
	p	HW	*p*	HW	*P*	HW	*p*	HW	*p*	HW	*p*	HW
Aacu45	0.000	0.619ns	0.000	1.000ns	0.290	0.030*	0.000	0.142ns	0.129	0.000*	0.148	0.000*
Aacu38	0.419	0.000*	0.326	0.000*	0.290	0.000*	0.290	0.000*	0.307	0.000*	0.258	0.000*
Aacu74[b]	–	–	0.000	0.123ns	0.290	0.142ns	0.210	0.440ns	0.611	0.000*	0.409	0.000*
Aac04	0.000	0.840ns	0.000	0.213ns	0.000	0.370ns	0.194	0.860ns	0.172	0.000*	0.101	0.000*
Aac12[b]	0.326	0.010*	–	–	0.290	0.040*	0.638	0.000*	0.056	0.025*	0.206	0.000*
sMo00137[b]	–	–	–	–	–	–	0.816	0.000*	0.991	0.000*	0.935	0.000*
sMo00020[b]	–	–	–	–	–	–	0.896	0.000*	0.995	0.000*	0.887	0.000*

PA=Pará; PE=Pernambuco; PB=Paraiba; SP=São Paulo; MG=Minas Gerais; MS=Mato Grosso do Sul; ns=not significant.
*significant deviation from HWE at α=0.05. [a] Frequency estimates (*p*) were based on the EM algorithm as described in the study by Dempster *et al*. (1977).
[b]No information (−) generated for loci with less than two alleles at the corresponding provenances.

4. DISCUSSION

The cross-amplification showed the ability of the polymorphic SSR markers to amplify the target microsatellite sequences in *A. aculeata*. The low percentage of the cross-amplification (26%) however could be attributed to the relatively wider taxonomic distance between the sources (*A. aculeatum* and *E. oleifera*) and the target (*A. aculeata*) species analyzed in the study, since all the species are from different genera (Baker *et al*., 2010). Nevertheless, when we look at the taxonomic relatedness between the species, both *A. aculeatum* and *A. aculeata* belong to the same sub-tribe (Bactridinae) and hence, in our investigation, provided a relatively higher percentage of cross-amplification (33%) as compared with *E. oleifera* (22%) (Table 3), which belongs to a different sub-tribe (Elaeidinae) (Baker *et al*., 2010).

As cited by Rosseto (2001), the success in the cross-species amplification of any DNA sequence is inversely related to the evolutionary distance between two species. According to Rossetto (2001), in plants there is high average rate of success in cross-species amplification of gSSR (genomic SSR) markers at sub-genus (89.8%) and genus (76.4%) levels than at family (35.2%) level. Using 20 of the SSR markers tested in this study (sourced from *E. oleifera* sets of SSRs), high cross-amplification percentage (83.3%) was reported when analyzed with *Elaeis guineensis* accessions (Zaki *et al*., 2012). Similarly, using the other 14 markers (sourced from *A. aculeatum* sets of SSRs), between 50 and 93% cross-amplification was reported in accessions of four *Astrocaryum* species (Ramos *et al*., 2012). These high cross-amplification percentages in the latter two studies were due to the fact that the cross-amplifications were at genus and species levels, respectively.

Despite of the low rate of cross-amplification obtained in the present study, the number of alleles amplified by the polymorphic markers was comparable with that in the study by Nucci *et al*. (2008), who first characterized polymorphic SSR markers for *A. aculeata*. Nucci *et al*. (2008) reported a range of two to eight alleles with an average of five per locus; while as already reported in this study, a range of three to nine alleles were detected by the markers with average of 5.4 per locus (Table 3). Out of the total number of alleles, a range of 2.4−6.8 effective numbers of alleles (N_e) were obtained in the provenances with an average of four per provenance (Table 3). According to Staub (1994), N_e represents measures of the number of equally frequent alleles, which are maintained in a population and used to make proper sampling in several generations. Similar results were obtained for heterozygosity and PIC when compared with the study by Nucci *et al*. (2008) (H_o = 0.27; H_E = 0.51 and PIC = 0.48) (Table 3).

However, the number of alleles reported for the four SSRs (Aac04, Aac12, sMo00020 and sMo00137) was higher in other studies when the markers were analyzed with accessions of *A. aculeatum* (average of six alleles per locus; Ramos *et al*. (2012)), and *E. oleifera* and *E. guineensis* (average of 7.5 alleles per locus; Zaki *et al*. (2012)). Similarly, average H_o, H_E and PIC per locus reported in Zaki *et al*. (2012) (H_o = 0.63; H_E = 0.82; PIC = 0.48) and Ramos *et al*. (2012) (H_o = 0.87; H_E = 0.76) were comparatively higher. As these markers were originally designed for the latter three species, a higher number of alleles, higher heterozygosity and PIC were expected than when analyzing with *A. aculeata*. These explain the species, genome and locus-specificity of the SSRs as already discussed by Powell *et al*. (1996) and Pinheiro *et al*. (2009); which increased the success of cross-amplification.

The average percentage of polymorphic loci (*P*=71.2 %) per provenance obtained in our study imply that SSRs are powerful molecular markers for assessing genetic diversity in *A. aculeata* populations. Besides, the markers used in this study are informative based on Bostein *et al*.'s (1980) classification of SSRs using PIC values (Table 3). Nevertheless, the polymorphism level reported here is still relatively lower than a RAPD (random amplified polymorphic DNA) polymorphism (79%) obtained in the same species (Oliveira *et al*., 2012). This could be explained by the lack of amplification products in some samples at certain SSR loci in the cross-amplification. Lack of amplification of an allele in certain accessions can be the result of null alleles potentially caused by PCR failure, the quality and quantity of DNA used, heterozygote deficits, scoring errors due to stuttering bands or other possible causes (Roa *et al*., 2000; Dakin and Avise, 2004). Hence, null alleles can no longer be detected by the primer, leading to an influence on the polymorphism level of SSR markers.

Our tests for the null allele showed evidence for the presence of null alleles across the loci (Table 4). The null allele frequency varied highly among the loci across the provenances demosntrating that the effects of the null alleles are locus specific (Dakin and Avise, 2004). As we observed from our results, loci with high heterozygote deficits showed strong null allele frequency than loci with less heterozygote deficits (Table 3 and Table 4). According to Chakraborty *et al*. (1992), inbreeding causes significant heterozygote deficits relative to HWE that might be misconstrued as evidence for null alleles. Our test for HWE showed significant deviation by most of the loci across the provenances analysed (Table 4). As discussed in the following paragraph, heterozygote deficits were detected in most of the loci analysed, which were potentially caused by inbreeding. Hence, when we compare the two results of null allele and HW tests, loci with a strong null allele frequency deviated significantly from HWE

owing to their heterozygote deficits, which are one of the potential causes of null alleles leading to low polymorphism (Dempster *et al*., 1977; Chakraborty *et al*., 1992).

Although *A. aculeata* has apparently mixed mating system (Scariot *et al*., 1991; Colombo *et al*., 2013), the monoecious nature of its inflorescence could favor more inbreeding, which resulted in reduced average heterozygosity per locus (Table 3). This was explained by positive inbreeding coefficients ($F > 0$) obtained across all loci (Table 3). Substantial positive inbreeding coefficient indicates the presence of inbreeding in a given population (Harrl and Clark, 1997). A case study on Jewelweed, *Impatiens capensis*, a common monoecious woodland flower in the Eastern US, demonstrated that the frequency of the homozygous genotypes increased with each generation of selfing, and the frequency of heterozygotes decreased (Stratton, 2008).

As described in the study by Zaki *et al*. (2012), amplification may not always represent functional SSRs for target species from cross-amplification. The PCR product of the markers need to be cloned and sequenced with selected individuals and the amplicon sequence need to be aligned with the original sequence from which the primers were designed. This may reveal the level of similarity and differences between the original gene sequence and the amplicon amplified by this specific locus. With this, markers can be identified based on sequence similarity in order to represent functional SSRs in the cross transferability. Therefore, it would be more appreciated if a further study will be conducted to confirm the inter-species transferability of the markers identified in this study by sequencing the amplicons and comparing with the original sequences from which the primers were designed.

5. CONCLUSIONS

We conclude that our investigation demonstrated the potential use of cross-species amplification to transfer SSR markers from two species of Arecaceae (*A.*

aculeatum and *E. oleifera*) to the newly emerging potential oil palm species, *A. aculeata.*

Along with the previously developed SSR markers (Nucci *et al*., 2008), the SSRs identified in the present study (Aac04, Aac12, sMo00020, sMo00137, Aacu38, Aacu45 and Aacu74) could provide invaluable support to realize various studies towards the domestication of *A. aculeata.* They could be used to study population genetics, germplasm characterization, genetic improvement and conservation.

The results also indicated that sourcing sets of SSRs from closely related taxa such as species or genus in Arecaceae could increase the success of transferring more number of functional markers. This would further facilitate comparative genetic studies between related species in Arecaceae, thus making the results directly comparative and the research more cost effective. Neverthless, for long term use and thorough study we recommend to invest more time and resource to construct genomic library for *A. aculeata* and develop new SSR markers.

6. REFERENCES

Abreu SI, Carvalho CR, Carvalho GMA and Motoike SY (2011) First karyotype, DNA C-value and AT/GC base composition of macaw palm (*Acrocomia aculeata*, Arecaceae) –a promising plant for biodiesel production. *Australian Journal of Botany* 59: 149–155.

Abreu AG, Priollig RH, Azevedo-Filho JA., Nucci SM, Zucchi SM, Coelho RM and Colombo CA (2012) The genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). *Genetics and Molecular Biology* 35: 119–121.

Baker WJ, Norup MV, Clarkson JJ, Couvreur TLP, Dowe JL, Lewis, CE, Pintaud, JC, Savolainen V, Wilmot T and Chase MW (2010) Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Annals of Botany* 108: 1417−1432.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF and Lexer C (2007) Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology* 16: 3759–3767.

Bora PS and Rocha RVM (2004) Macaíba palm: fatty and amino acids composition of fruits. *Ciência e Tecnologia de Alimentos* 4: 158–162.

Bostein D, Whiter L, Skolnick M and Davis RW (1980) Construction of a Genetic Linkage Map in Man using Restriction Fragment Length Polymorphisms. *The American Journal of Human Genetics* 32: 314–331.

Brito GG, Caixeta ET, Gallina PA, Zambolim EM, Zambolim L, Diola V and Loureiro ME (2010) Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* 173: 255–264.

Chakraborty R, De Andrade M, Daiger SP and Budowle B (1992) Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* 56: 45–57.

Chapuis MP and Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Cole CT (2003) Genetic variation in rare and common plants. *Annual Review of Ecology, Evolution and Systematics* 34: 213−237.

Colombo CA, Nucci SM, Priolli RHG, Zucchi MI, Carhalho CRL, Chorfl BLH, Siqueira WG and Azvedo-Filho JÁ (2013) Genetic Diversity of Macaw Palm (*Acrocomia aculeata*) by Microsatellite Markers. International Plant and Animal Genome Conference. Available at http://www.intlpag.org.

Cruz CD (2013) GENES V9.0. *Software package for analysis in quantitative genetics and experimental statistics*, Universidade Federal de Viços, Brazil.

Cruz CD, Ferreira FM and Pessoni LA (2011) Biometria aplicada ao estudo da diversidade genética.Viçosa: Editora Universidade Federal de Viçosa, p.622.

Dakin EE and Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity* 93: 504−509.

Dempster AP, Laird NM and Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of The Royal Statistical Society* 39:1–38.

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

Faleiro FG, Costa AM, Karia CT, Andrade RP, Junqueira NTV, Pereira AV, Pereira EB and Sano SM (2008) Molecular markers and geographic information systems as a tool to study native plant species in the Brazilian Savannas. In: *The 2nd International Symposium in Tropical Savannas*. 12−17 October 2008, ParlaMundi, Brazilia, available at http://www.cpac.embrapa.br.

Fortes ICP and Baugh PJ (1999) Study of analytical on-line pyrolysis of oils from Macauba fruit (*Acrocomia sclerocarpa* M.) via GC/MS. *Journal of The Brazilian Chemical Society* 10: 469–477.

Hartl DL and Clark AG (1997) *Principles of population genetics*. USA: Sinauer Associates, p.542.

Hiane PA, Ramos-Filho MM, Ramos MIL and Macedo MLR (2005) Bocaiúva, *Acrocomia aculeata* (Jacq.) Lodd., pulp and kernel oils: characterization and fatty acid composition. *Brazilian Journal of Food Technology* 8: 256–259.

Kantety RV, Rota ML, Matthews DE and Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* 48: 501–510.

Lanes ECM, Nick C, Kuki KN, Freitas RD and Motoike SY (2013) Genomic DNA isolation of *Acrocomia aculeata* (Arecaceae) from leaf and stipe tissue samples for PCR analysis. *Genetics and Molecular Research* 12: 3905-3911.

Lanes ECM, Costa PMA and Motoike SY (2014) Alternative fuels: Brazil promotes aviation biofuels. *Nature* 511:31

Motoike SY and Kuki KN (2009) The potential of macaw palm (*Acrocomia aculeata*) as source of biodiesel in Brazil. *International Review of Chemical Engineering* 1: 632–635.

Motoike SY, Lopes FA, Sá Júnior AQ, Carvalho M and Oliveira MAR (2007) Processo de Germinação e Produção de Sementes Pré-Germinadas de Palmeiras do Gênero *Acrocomia*. Submetido à Lei de Patentes. Protocolo INPI: 014070005335.

Moura EF, Motoike SY, Ventrella MC, Sá Junior AQ and Carvalho M (2009) Somatic embryogenesis in macaw palm (*Acrocomia aculeata*) from zygotic embryos. *Scientia Horticulturae* 119: 447–454.

Nei M (1978) Estimation of average heterozygosity and genetic distances from a small number of individuals. *Genetics* 89: 583–590.

Nucci SM (2007) Development, characterization and analyses of the use of microsatellite markers in macaw palm population genetics. MSc. Thesis, University of Sâo Paulo.

Nucci SM, Azevedo-Filho A, Colomo AC, Priolli GHR, Coelho MR, Mata TL and Zucchi IM (2008) Development and characterization of microsatellites markers from the macaw. *Molecular Ecology Resources* 8: 224–226.

Oliveira DA, Júnior-Melo AF, Brandão MM, Rodrigues LA, Menezes EV and Ferreira PRB (2012) Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais. *Genetics and Molecular Research* 11: 531–538.

Pinheiro F, Palma-Silva C, Barros F and Cozzolino S (2009) Cross-amplification and characterization of microsatellite loci for the Neotropical orchid genus *Epidendrum*. *Genetics and Molecular Biology* 32: 337–339.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Moleculur Breeding* 2: 225–238.

Ramos SF, Vasconcelos de Macêdo JL, Lopes MTG, Batista JS, Formiga KM, Pimentel da Silva P, Saulo-Machado AC and Veassey EA (2012) Microsatellite Loci for Tucumã of Amazonas (*Astrocaryum aculeatum*) and amplification in other Arecaceae. *American Journal of Botany* 99: e508-e510.

Raymond M and Rousset F (1995) GENEPOP V1.2. Population genetics software for exact tests and ecumenicism. *Heredity* 86:248–249.

Ribeiro LM, Souza PP, Rodrigues AG, Oliveira TGS and Garcia QS (2011) Overcoming dormancy in macaw palm diaspores, a tropical species with potential for use as bio-fuel. *Seed Science and Technology* 39: 303–317.

Roa AC, Chavarriaga-Aquire P, Dugue MC, Maya MM, Bonierbale MW, Iglesias C and Tohme JM (2000) Cross species-amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationships. *American Journal of Botany* 87: 1647–1655.

Rossetto M (2001) Sourcing of SSR markers from related plant species. In: Henry RG (ed.) *Plant Genotyping: The DNA finger printing of plants*. New York. CAB International, pp. 211–224.

Scariot AO, Lleras E and Hay JD (1991) Reproductive biology of the palm *Acrocomia aculeata* in Central Brazil. *Biotropica* 23: 12–22.

Staub JE (1994) Evolution and population. In: Crossover: Concepts and Applications in Genetics, Evolution, and Breeding. An interactive computer- based laboratory manual. USA: University of Wisconsin Press, p. 359.

Stratton D (2008) Case Studies in Ecology and Evolution (DRAFT). Non-random mating, Inbreeding and Population Structure, available at http://www.uvm.edu.

Teixeira LC (2005) Potencialidades de oleaginosas para produção de biodiesel. In: Lacerda V (ed.) *Produção de Oleaginosas para Bioiesel*. Belo Horizonte: Informe Agropecuário, pp 18–27.

Tickel J (2000) *From the Fryer to the Fuel Tank: The Complete Guide to using Vegetable Oil as an Alternative Fuel*. Florida: Tickel Energy Consulting, p.155.

Zaki NM, Singh R, Rosli R and Ismail I (2012) *Elaeis oleifera* Genomic-SSR Markers: Exploitation in Oil Palm Germplasm Diversity and Cross-Amplification in Arecaceae. *International Journal of Molecular Sciences* 13: 469−4088.

CHAPTER 2

GENETIC DIVERSITY IN THE MACAW PALM (*Acrocomia aculeata*) GERMPLASM ACCESSIONS BASED ON MICROSATELLITE MARKERS.

ABSTRACT

Macaw palm (*Acrocomia aculeata*) is native to tropical forests in South America, widely distributed and abundant in Brazil. It is highly productive oleaginous palm tree presenting high potential for biodiesel production. In spite of its immense potential as a biodiesel plant, *A. aculeata* is still in the wild and underutilized. Besides, its genetic resources are threatened by loss of natural habitat, predatory extractivism, unsustainable land use and climate change. Efforts are underway in conserving its genetic resources *ex situ* to ensure for the launching of macaw palm breeding program. Hence, generating knowledge on genetic diversity of the resources is unprecedented to support for planning efficient collection and sustainable conservation strategies. The aim of this work was to study the genetic diversity of *A. aculeata ex situ* germplasm collections using microsatellite (Simple Sequence Repeats-SSR) markers. A total of 192 accessions of 41 different families of *A. aculeata*, originally collected from six provenances in Brazil were analyzed with 10 polymorphic SSR markers. A total of 72 alleles with an average of 7.2 per locus were obtained. Consequently high percentage of polymorphic loci (92−100%) was detected in the provenances revealing variation in allelic diversity, heterozygosity and polymorphism in the collections analyzed. Reduced average observed heterozygosity (H_o<50%) per locus evidenced heterozygote deficiency in the germplasm collections and this was confirmed by a positive inbreeding coefficient (F > 0) detected per locus (or provenance). Based on genetic distances, three distinct groups of accessions were formed using: i) unweighted pair-group method with arithmetic means (UPGMA); ii) optimization method of Tocher and iii) graphical dispersion using principal coordinate analysis (PCoA) and projection with 3D principal components. Nevertheless, a non-significant correlation (r=0.4128) was obtained between genetic and geographic distances of the provenances evidencing genetic similarities among accessions of distant provenances that proved the dispersal of *A. aculeata* over long distances in Brazil. Analysis of Molecular Variance (AMOVA) revealed higher (48.22%) genetic variation within family than among families (36.47%) followed by variation among provenances (15.31%). The study demonstrated that there exists

genetic variability in the *ex situ* collections of *A. aculeata* in the germplasm bank, in which the information could be utilized in the pre-breeding program. The study also showed the efficiency of SSR markers in germplasm characterization in *A. aculeata* and the information generated could be used in formulating appropriate collection and conservation strategies.

Keywords: biodiesel; conservation; domestication; germplasm bank; macaúba; SSR

DIVERSIDADE GENÉTICA EM ACESSOS DE GERMOPLASMA DE MACAÚBA (*Acrocomia aculeata*) COM BASE EM MARCADORES MICROSSATÉLITES.

RESUMO

A palmeira macaúba (*Acrocomia aculeata*) é nativa das florestas tropicais da América do Sul, amplamente distribuída e abundante no Brasil. É uma palmeira oleaginosa, altamente produtiva, que apresenta grande potencial para produção de biodiesel. Apesar de seu imenso potencial como uma planta de biodiesel, *A. aculeata* ainda está em meio natural e subutilizados. Além disso, seus recursos genéticos são ameaçadas pela perda de habitat natural, o extrativismo predatório, uso insustentável da terra e mudanças climáticas. Esforços estão sendo na conservação de seus recursos genéticos *ex situ* para assegurar para o lançamento do programa de melhoramento de macaúba. Assim, informações sobre variabilidade genética em coleções de germoplasma precisam ser gerados de forma a apoiar a planejamento de programa de melhoramento. O objetivo deste trabalho foi estudar a diversidade genética das coleções *ex situ* de germoplasma da *A. aculeata* utilizando microssatélites (Simple Sequence Repete-SSR). Um total de 192 acessos de 41 famílias de seis procedências foram analisados com 10 marcadores SSR. Na análise, foi obtido um total de 72 alelos, com média de 7,2 por loco. Consequentemente, observou-se alta porcentagem de locos polimórficos (92−100%) entre as procedências, revelando variação na diversidade alélica, heterozigosidade e polimorfismo nos acessos analisados. A heterozigosidade por loco observada foi reduzida (Ho<50%), evidenciado deficiência de heterozigotos nas coleções de germoplasma, e que foi confirmado pelo coeficiente de endogamia positivo (F > 0) por loco (ou de procedência). Baseado nas distâncias genéticas, três grupos distintos de acessos foram formados utilizando: i) método das médias aritméticas não ponderadas (UPGMA); ii) método de otimização de Tocher e iii) a dispersão geográfica usando analise de coordinadas principais (PCoA) e projeção em componentes principais 3D. No entanto, uma correlação não-significativa (r=0,4128) foi obtido entre distâncias genéticas e geográficas detectando semelhanças genéticas entre procedências distantes que provaram a dispersão de *A. aculeata* no Brasil em distâncias mais longas. A análise de variância molecular (AMOVA) revelou maior variação genética (48,22%) dentro da família do que entre as famílias (36,47%), seguido de variação entre as procedências (15,31%). O estudo demonstrou, que existe variabilidade genética nas coleções *ex situ* do banco de germoplasma, em que a informação poderia ser utilizada no programa de

pré-melhoramento. O estudo também mostrou a eficiência dos marcadores SSR na caracterização de germoplasma de *A. aculeata* e as informações geradas podem ser utilizados na formulação de estratégias adequadas de coleção e de conservação.

Palavras-chave: banco de germoplasma; biodiesel; conservação; domesticação; macaúba; SSR.

1. INTRODUCTION

Macaw palm (*Acrocomia aculeata* (Jack.) Lodd. ex Mart. 2n=2x=30) belongs to the family of Arecaceae; arborescent, spiny and single-stemmed palm (Read, 1966; Abreu *et al*., 2011). It is monoecious experiencing entomophily and anemophily forms of pollinations which bear mixed reproductive systems (Scariot *et al*., 1991, 1995; Abreu *et al*., 2012). *A. aculeata* has large and wide distribution in the dry regions of tropical America and it is native palm tree of tropical forests, occurring from southern Mexico to southern Brazil, Paraguay and Argentina (Henderson *et al*., 1995; Morcote-Rios and Bernal, 2001). In Brazil, it has abundant distribution in the regional States of Ceará, Minas Gerais, Mato Grosso, Mato Grosso do Sul and São Paulo (Scariot *et al*., 1991; Henderson *et al*., 1995;).

A. aculeata is little known globally, however in recent years, high interest has aroused due to its potential for social and economic use as an oil producer, considering that it is cited as one of the most important new sources of oil to make biofuel for both aviation and automotive sectors (Nucci *et al*., 2008; Lanes *et al*., 2014). It produces fruits up to 25 tonnes/ha which could be processed to 4000 kg of oil. The solid waste is converted to charcoal and nutritious cakes that can be used to generate energy and feed livestock (Tickel, 2000; Moura *et al*., 2009). The biochemical properties of the oil are proved to be suitable for the cosmetic industry and for biodiesel production (Fortes and Baugh, 1999; Bora and Rocha, 2004; Hiane *et al*., 2005). Moreover, this palm has environmental benefits as it can foster in impoverished soils and drought prevailing areas, which is a desirable trait for plants to rehabilitate degraded pastures or for agroforestry practices (Motoike and Kuki, 2009).

Although it has not been cultivated commercially like the domesticated Arecaceae palms such as *Elaeis guineensis*, *Cocus nucifera* and *Enterpe oleraceae*, as an important element in the Brazilian savannah, *A. aculeata* possess great genetic

diversity in natural populations (Nucci *et al*., 2008; Oliveira *et al*., 2012; Colombo *et al*., 2013). Nevertheless, the natural genetic resource of this important palm has been threatened by many factors such as loss of natural habitat, predatory extractivism, unsustainable land use and climate change (Faleiro *et al*., 2008; Ribeiro *et al*., 2011). Hence, conserving the genetic resources of this important palm will have paramount importance to help for future genetic improvement and sustainable use of its genetic resources.

Central to sustainable conservation is the knowledge of the genetic diversity present in a gene bank collections and potential exploitation of the genetic materials by breeding programs. Germplasm characterizations using SSR markers are well known for their potentially high information content and versatility as molecular tools (Ferreira, 2005) detecting genetic variations and polymorphic traits without influence of the environment (Tanksley, 1989). Hence, germplasm characterization in *A. aculeata* using SSR markers can have invaluable contribution for the establishment of breeding programs and conservation of genetic resources.

Therefore, this study was conducted to study the extent of genetic diversity in macaw palm *ex situ* germplasm accessions using SSR markers.

2. MATERIALS AND METHODS

2.1 Plant material and DNA isolation

Leaf samples from 192 accessions of *A. aculeata* were obtained from *ex situ* plant collection, macaúba Active Germplasm Bank (BAG – Macaúba respository nº: 084/2013/CGEN/MMA) located in the experimental farm of the Universidade Federal de Viçosa in the municipality of Araponga ($20^{o}40'1''$S, $42^{o}31'15''$W), State of Minas Gerais, Brazil (Figure 1).

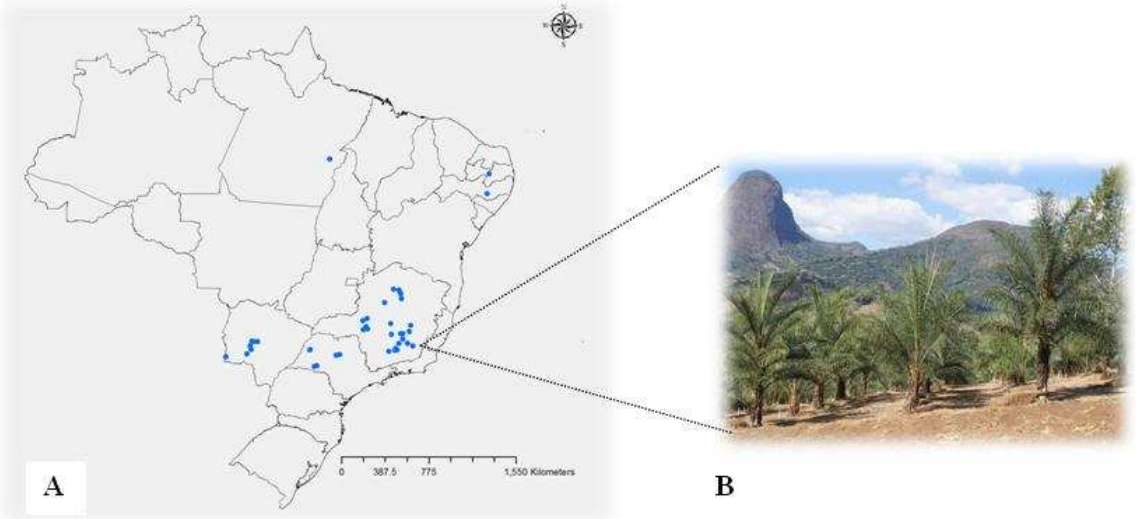

Figure 1. (A) Map shows the six provenances, in which the families were originally collected; (B) part of the experimental field in Araponga-Minas Gerais, where the germplasm bank (BAG-Macaúba) is located.

The accessions were originally developed from seeds collected from six provenances in Brazil (Figure 1A) and germinated using a pre-germination protocol as described in patent INPI 014070005335 (Motoike *et al*., 2007). The accessions represent individuals of a total of 41 different families coded as BGP. Each accession has three to five individuals included in the study (Table 1).

Genomic DNA (gDNA) was isolated from young leaf of each individual sample using Cetyl trimethylamonium bromide (CTAB) method (Doyle and Doyle, 1990) with modifications as described by Lanes *et al*. (2013).

DNA samples were quantified with Multiscan^{TM} GO/UV Microplate Spectrophotometer (Thermo Fisher Scientific OY, Ratasite, Finland). A 2µl of DNA for each sample was added into a 96 well-plate, which was mounted on the spectrophotometer and absorbance reading was taken at 260 and 280nm. A ratio of 260/280 nm absorbance was calculated and only values within the range of 1.8 to 2.0 were considered as it denotes the absorption is due to nucleic acids (Hoisington *et al*., 1994). DNA samples with absorbance ratio < 1.8 or > 2.0 were not considered as it

denotes contamination with proteins or phenol compounds; hence DNA extraction was repeated for these samples.

$$DNA\ concentration\ (\mu g/ml) = Absorption_{260nm} \times 50 \mu g/ml$$

1 absorbance at 260nm corresponds to a concentration of 50µg/ml for double-stranded DNA (standard)

The integrity of the DNA samples was confirmed on 2% agaros gel electrophoresis and the working concentration was adjusted to 30ng/µl for PCR.

Table 1. List of 41 *Acrocomia aculeata* families, number of individuals per family, provenance and GPS coordinates of each collection point

No.	Family code	Number of individuals	Provenance*	Coordinates**		No.	Family code	Number of individuals	Provenance*	Coordinates**	
				Latitude	Longitude					Latitude	Longitude
1	BGP99	5	PA	S 06 03 58.0	W 49 33 39.0	22	BGP11	5	EMG	S 19 14 01.2	W 43 03 28.4
2	BGP82	5	PE	S 07 14 23.0	W 36 46 55.0	23	BGP9	5	EMG	S 19 33 12.0	W 46 51 10.1
3	BGP124	4	PB	S 08 48 49.0	W 36 57 14.0	24	BGP78	5	EMG	S 18 51 25.6	W 46 52 55.2
4	BGP51	5	SP	S 21 32 04.6	W 48 44 24.7	25	BGP37	5	EMG	S 18 40 51.3	W 46 33 41.4
5	BGP34	5	SP	S 22 25 10.8	W 50 34 43.1	26	BGP33	5	EMG	S 19 19 40.3	W 46 38 11.5
6	BGP47	5	SP	S 22 29 14.2	W 50 46 16.2	27	BGP21	4	WMG	S 19 31 15.9	W 46 31 42.2
7	BGP20	5	NMG	S 16 39 52.7	W 43 53 58.9	28	BGP2	5	WMG	S 20 39 20.4	W 43 18 45.2
8	BGP27	5	NMG	S 16 21 20.7	W 44 25 30.5	29	BGP76	5	WMG	S 19 41 51.4	W 43 11 27.7
9	BGP22	4	NMG	S 17 25 54.0	W 45 08 59.5	30	BGP25	5	WMG	S 17 06 54.6	W 43 49 16.4
10	BGP16	5	NMG	S 16 26 07.6	W 44 00 50.5	31	BGP64	5	WMG	S 16 44 12.7	W 43 51 54.9
11	BGP49	5	NMG	S 20 38 58.0	W 44 01 15.5	32	BGP105	3	MS	S 20 29 52.5	W 55 18 39.3
12	BGP10	5	NMG	S 21 03 12.9	W 44 16 28.2	33	BGP102	4	MS	S 20 30 38.6	W 55 37 59.7
13	BGP68	5	SMG	S 21 11 27.6	W 44 19 29.7	34	BGP104	4	MS	S 20 27 55.9	W 55 46 41.7
14	BGP3	5	SMG	S 21 09 52.2	W 44 08 49.5	35	BGP117	3	MS	S 20 27 56.5	W 55 46 38.2
15	BGP51	5	SMG	S 21 17 20.5	W 44 49 12.6	36	BGP118	4	MS	S 20 50 22.3	W 55 54 53.3
16	BGP5	5	SMG	S 19 05 02.0	W 44 39 13.9	37	BGP112	5	MS	S 20 50 16.5	W 55 54 51.8
17	BGP14	5	SMG	S 19 56 29.0	W 44 36 12.0	38	BGP106	3	MS	S 21 28 42.3	W 56 10 03.6
18	BGP18	5	CMG	S 19 52 34.0	W 43 52 20.5	39	BGP92	5	MS	S 21 28 45.7	W 56 10 06.6
19	BGP24	5	CMG	S 19 53 20.2	W 43 41 11.5	40	BGP103	5	MS	S 21 42 04.8	W 57 50 39.0
20	BGP1	5	CMG	S 20 17 42.6	W 43 42 30.9	41	BGP119	4	MS	S 21 42 06.0	W 57 50 32.4
21	BGP52	5	CMG	S 20 50 13.1	W 42 54 27.3						

*Provenances: PA=Pará, PE=Pernambuco, PB=Paraiba, SP=São Paulo, MG=Minas Gerais containing five sub-provenances (NMG=North Minas Gerais, SMG=South Minas Gerais, CMG=Central Minas Gerais, EMG=East Minas Gerais, and WMG=West Minas Gerais) and MS=Mato Grosso do Sul.

**Coordinates are in degrees, minutes and seconds for both the latitude (S=South) and longitude (W=West).

2.2 Condition of Polymerase Chain Reaction (PCR) and Electrophoresis

The PCR was performed according to Nucci *et al*. (2008) with some modifications. It was carried out in a 20µL total reaction volume containing 30ng of *A. aculeata* genomic DNA; 1xPCR buffer; 250µM of each dNTP; 0.15µM of forward and reverse primer; 3mM of $MgCl_2$ (4mM of $MgCl_2$ was used for primer Aac04 and Aac12 (Table 2) for better amplification) and 1U of *Taq* DNA polymerase (invitrogen). The amplification cycles were carried out in a thermal cycler (Applied Biosystem®Verti®cycler, Thermo Fisher Scientific Brand, USA) programmed at initial denaturing step at 94^{o}C for 5min followed by 30 cycles of 1min denaturation at 94^{o}C; annealing at primer-specific annealing temperature for 1min (Table 2); extension at 72^{o}C for 1min and the final extension at 72^{o}C for 8min.

A total of 10 SSR primers were used, of which five from Nucci *et al*. (2008); three from Nucci (2007) and the other two were obtained from sets of SSR makers originally developed for *Astrocaryum aculeatum* (Ramos *et al*., 2012) and selected for *A. aculeata* through transferability (Mengistu *et al*., 2015) (Table 2). PCR products were denatured in a bromophenol blue dye solution at 95^{o}C for 5min in the thermal cycler just before running on 6% denaturing polyacrylamide gel electrophoresis in 1xTBE (Tris-Borate-EDTA) buffer solution at 60W for 1hr and 40min.

2.3 Polyacrylamide gel staining

After electrophoresis, the PCR products were visualized in polyacrylamide gels stained with silver nitrate ($AgNO_3$) solution according to Brito *et al*. (2010). The stained gels were allowed to dry in the air and DNA fragments were scored as co-dominant alleles.

2.4 Data analyses

Allelic diversity, heterozygosity and polymorphism levels of the SSR markers and provenances were estimated. Different methods of grouping were applied to study the extent of genetic diversity in *A. aculeata* accessions. The parameters analyzed and the analytical methods used were:

i) Number of alleles per locus (*A*), determined by quantifying the number of different alleles detected per locus. Hence, total number of alleles per provenance (N_t) determined by summing up the number of alleles amplified by each locus: $N_t = \sum_{j=1}^{a} a_j$, where *a* is the number of alleles at j^{th} locus. Consequently, average number of alleles per provenance was determined as $N_a = \frac{1}{L}\sum_{j}^{a} a_j$; where *L* is the number of loci analyzed (Cruz *et al*., 2011).

ii) Observed heterozygosity per locus was calculated as H_o= (number of heterozygotes observed/ total number of individuals analyzed); hence average observed heterozygosity per provenance was estimated as $\bar{H}_o = \frac{1}{L}\sum_{j=1}^{L} H_{o(j)}$; where H_o is observed heterozygosity at j^{th} locus and *L* is number of loci analyzed. Expected heterozygosity per locus was calculated as $H_E = 1 - \sum_{j=1}^{a} p_i^2$ where *pi* is the frequency of the i^{th} allele at j^{th} locus (Nei, 1978).

iii) Inbreeding coefficient per locus (or provenance) was estimated to determine deficiency (or excess of heterozygosity) as $F = 1 - \frac{Ho}{HE}$; where H_o and H_E are average observed and expected heterozygosity per locus respectively (Hartl and Clark, 1997).

iv) Polymorphic information content per locus was calculated as $PIC = 1 - \sum_{j=1}^{a} p_i^2 - \sum_{i,j=1}^{a}\sum_{(i\#j)}^{a} p_i^2\, p_j^2$; where *pi* and *pj* are frequencies of the i^{th} allele at j^{th} locus (Bostein *et al*., 1980).

v) Percentage of polymorphic loci per provenance was calculated as *P* = (number of polymorphic loci/ total number of loci analyzed) based on three criteria set

by Cole (2003): (a) when a locus detected polymorphism in at least one individual of the sample; (b) when loci had common alleles with a frequency of less than 0.99, or (c) when loci had common alleles with a frequency of less than 0.95. These criteria were designated herein as *P*100, *P*99, and *P*95, respectively.

vi) Genetic distance between pairs of provenances was computed using the formula: $Nei\_D = -\ln(I)$. Where $I = \frac{J_{xy}}{\sqrt{J_x J_y}}$, in which; I is Nei's Genetic Identity. $J_x = \frac{1}{L}\sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijx}^2$; $J_y = \frac{1}{L}\sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijy}^2$ and $J_{xy} = \frac{1}{L}\sum_{j=1}^{L}\sum_{k=1}^{a_j} p_{ijx}\, p_{ijy}$; where p_{ijx} and p_{ijy} are frequencies of the i^{th} allele at j^{th} locus of provenance *x* and *y* respectively and *L* stands for the number of loci (Nei, 1972).

vii) Cluster analysis was applied to construct dendogram from Nei's Genetic distance matrix between pairs of provenances using unweighted pair-group method with arithmetic means (UPGMA). Tocher was also applied to form homogenous groups of provenances from Nei's genetic distance matrix.

viii) Principal coordinate analysis (PCoA) was applied for graphical dispersion of the accessions on bi-dimensional axes. PCoA was done from the genetic distance between pairs of accessions according to Peakall and Smouse (2006).

ix) Mantel test using Pearson correlation was performed to test for the relationships between genetic and geographic distances among *A. aculeata* provenances analyzed from different regional States in Brazil. Geographic distance (km) among the provenances was calculated using Geographic distance matrix generator (Ersts, 2014).

x) Analysis of Molecular Variance (AMOVA) was done to estimate the amount of genetic variance in the accessions analayzed. Components of variances were estimated to determine the percentage of molecular variance for each source of variation and Φ-statistics were computed from the variances to test for the null hypothesis: H_0: $\sigma_a^2 = \sigma_b^2 = \sigma_c^2 = 0$; where σ_a^2 =variance among provenances; σ_b^2 =variance among

families; and σ_c^2 =variance within family. The Φ-statistics were calculated as $\Phi_{GT} = \frac{\hat{\sigma}_a^2}{\hat{\sigma}_T^2}$; $\Phi_{SG} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2+\hat{\sigma}_c^2}$ and $\Phi_{ST} = \frac{\hat{\sigma}_a^2+\hat{\sigma}_b^2}{\hat{\sigma}_T^2}$ (Excoffier *et al*., 1992; Excoffier, 2001). The calculated Φ-values were compared for significance tests with values obtained under 1000 permutations. Data analyses were performed using GENES (Cruz, 2013) and GenAIEx (Peakall and Smouse, 2006) statistical softwares.

3. RESULTS AND DISCUSSION

3.1 Allelic polymorphism and heterozygosity

A total of 72 alleles were detected in the analysis of 192 *A. aculeata* individuals (or accessions) using ten SSR markers. A range of 4 (Aac12) to 11 (Aacu12) alleles were obtained with average of 7.2 per locus (Table 2).

In similar study by Nucci *et al*. (2008), a total of 30 alleles with average of 6 per locus were reported in the analysis of *A. aculeata* accessions from São Paulo and Minas Gerais populations using five of the SSR markers (Aacu07, Aacu10, Aacu12, Aacu26 and Aacu30). In this study however, these five SSR produced 40 alleles with average of 8 per locus; which is relatively higher as compared to Nucci *et al*. (2008) (Table 2).

The wider coverage of the genetic materials analyzed from the six provenances in the present study led to yield more number of alleles per locus and consequently resulted in more allelic diversity.

Table 2. Primer pairs of 10 SSR markers used in the study and average values obtained for different parameters per locus.

Locus	Forward and Reverse primer sequence (5´-3´)	Allele size (bp)	*A*	H_o	H_E	F	PIC	T_a (^oC)
Aacu07[a]	F: ATCGAAGGCCCTCCAATACT R: AAATAAGGGGACCCTCCAA	153-177	6	0.43	0.48	0.10	0.43	56
Aacu10[a]	F: TGCCACATAGAGTGCTTGCT R: CTACCACATCCCCGTGAGTT	168-186	8	0.58	0.69	0.16	0.65	56
Aacu12[a]	F: GAATGTGCGTGCTCAAAATG R: AATGCCAAGTGACCAAGTCC	190-202	11	0.57	0.71	0.20	0.67	56
Aacu26[a]	F: ACTTGCAGCCCCATATTCAG R: CAGGAACAGAGGCAAGTTC	273-316	9	0.41	0.63	0.35	0.56	56
Aacu30[a]	F: TGTGGAAGAAACAGGTCCC R: TCGCCTTGAGAAATTATGGC	148-158	6	0.39	0.43	0.09	0.38	56
Aacu38[b]	F: TTCTCAGTTTCGTGCGTGAG R: GGGAGGCATGAGGAATACAA	316-346	6	0.13	0.64	0.80	0.58	56
Aacu45[b]	F: CAGACTACCAGGCTTCCAGC R: TCATCATCGCAGCTTGACTC	260-284	5	0.30	0.38	0.21	0.34	56
Aacu74[b]	F: TACTGTTGTGCCAAGTCCCA R: GAGCACAAGGGGGATATCAA	278-313	9	0.26	0.45	0.42	0.42	56
Aac04[c]	F: GCATTGTCATCTGCAACCAC R: GCAGGGGCCATAAGTCATAA	258-306	8	0.61	0.72	0.15	0.68	60
Aac12[c]	F: GCTCTGTAATCTCGGCTTCCT R: TCCAGTTCAAGCTCTCTCAGC	229-247	4	0.06	0.31	0.81	0.27	60
Mean		-	7.2	0.37	0.54	0.33	0.50	-

a, Nucci *et al*. (2008); b, Nucci (2007); c, Mengistu *et al*. (2015); *A*=number of alleles per locus; H_o = observed heterozygosity per locus; H_E=expected heterozygosity per locus; F= inbreeding coefficient per locus; PIC=polymorphic information content per locus; T_a= primer annealing temperature.

The variation in frequency of genotypes resulted in variation in average observed heterozygosity (H_o) among the loci. H_o ranged from 0.06 (Aac12) to 0.61(Aac04) with average of 0.37 per locus (Table 2). The low proportion of H_o per locus (< 50%) could be caused by inbreeding or crossing between genetically related parents yielding the progenies (or families) which were analyzed in this study. This was confirmed by a positive inbreeding coefficient (F > 0) with average of 0.33 per locus (Table 2). According to Harrl and Clark (1997), a substantial positive inbreeding coefficient per locus or population indicates the predominance involvement of inbreeding. Although *A. aculeata* has mixed mating system (Scariot *et al*., 1991; Abreu *et al*., 2012; Colombo *et al*., 2013), the monoecious (androgenous) nature of its inflorescence marked with protogyny favors pollination between different flowers on

the same plant (geitonogamy) or between plants (xenogamy). However, according to Scariot *et al*. (1991), geitonogamy is highly responsible for the majority of the fruit set in *A. aculeata* than xenogamy, which could potentially reduce the proportion of heterozygotes in the progenies. Other factors, for instance: gregarious nature of the species and fragmentation of habitats leading to formation of isolated populations, which could suffer from genetic drift and restricted gene flow contributing to endogamy.

Polymorphic information content (PIC) ranged from 0.27 (Aac12) to 0.68 (Aac04) with average of 0.50 per locus (Table 2). According to Bostein *et al*.'s (1980), classification of SSR markers, all markers analyzed are quite informative and polymorphic to characterize the germplasm accessions (or populations) in *A. aculeata.* SSR markers are classified as very informative when PIC > 0.50, reasonably informative (0.25< PIC < 0.50) or less informative (PIC < 0.25).

3.2 Genetic diversity

High average percentage of polymorphic loci (*P*) per provenance was obtained with a range of 92 to 100% based on the three criteria proposed by Cole (2003) (Table 3). The total number of alleles (N_t) detected per provenance varied from 28 (Paraiba) to 57 (Mato Grosso do Sul). The average number of alleles per provenance (N_a) ranged from 2.8 to 5.7 respectively (Table 3). Hence, the highest proportion of alleles (78%) was obtained in Mato Grosso do Sul and the least (38%) in Paraiba. The variation in proportion of alleles (or allelic diversity) and the high percentages of polymorphic loci among the provenances attested the presence of genetically diverse germplasm collections in the germplasm bank.

With the first criteria proposed by Cole (2003), when a locus detected polymorphism in at least one individual of the sample, 100% *P* was obtained in all the

provenances. However, with the second and the third criteria, 93% and 92% average *P* was obtained per provenance with alleles of a frequency of less than 0.99 and 0.95 respectively (Table 3).

Table 3. Average estimates of some genetic diversity parameters for *Acrocomia aculeata* provenances based on ten polymorphic SSR markers.

						P (%)		
Provenances	N_t	P_a	N_a	$\bar{H}_o$	*F*	P_{100}	P_{99}	P_{95}
PA	32	0.44	3.20	0.46	0.15	100	90	90
PE	29	0.40	2.90	0.34	0.24	100	90	90
PB	28	0.38	2.80	0.35	0.29	100	100	100
SP	39	0.53	3.90	0.32	0.42	100	100	90
NMG	52	0.71	5.20	0.32	0.48	100	90	90
SMG	39	0.53	3.90	0.36	0.32	100	80	80
CMG	42	0.58	4.20	0.29	0.48	100	90	90
EMG	43	0.59	4.30	0.30	0.40	100	90	90
WMG	55	0.75	5.10	0.28	0.55	100	100	100
MS	57	0.78	5.70	0.46	0.22	100	100	100
Mean	42	-	4.12	0.35	0.36	100	93	92

N_t=total number of alleles per provenance; P_a=proportion of alleles per provenance; N_a=average number of alleles per provenance; $\bar{H}_o$= average observed heterozygosity per provenance; F= inbreeding coefficient per provenance; *P*(%)=percentage of polymorphic loci with three criteria (P_{100}, P_{99} and P_{95}). PA=Pará, PE=Pernambuco, PB=Paraiba, SP=São Paulo, NMG=North Minas Gerais, SMG=South Minas Gerais, CMG=Central Minas Gerais, EMG=East Minas Gerais, WMG=West Minas Gerais and MS=Mato Grosso do Sul.

The *P* values obtained (92−100%) using the three criteria are much higher when compared with the result obtained by Oliveira *et al*. (2012) (*P*=79%), who studied *A. aculeata* from natural populations using RAPD markers. This could be explained by the high polymorphic level of SSR as compared to RAPD markers. Hence, the results attested that, the SSR markers used in this study are polymorphic molecular markers to study diversity in *A. aculeata.* Moreover, the reduced average observed heterozygosity ($\bar{H}_o$< 50%) and positive inbreeding coefficient (F > 0) obtained at provenance level (Table 3) demonstrated a similar situation as already discussed above in the allelic polymorphism and heterozygosity part at locus level (Table 2). It reaffirmed that the

accession studied were resulted from either endogamy or crossing among genetically related individuals which potentially reduced the proportion of heterozygotes.

Based on genetic distance analyses, similar grouping of accessions were obtained using three different methods of grouping. Using UPGMA, three hierarchical groups of *A. aculeata* provenances were established when a line cuts at distance of 0.66, corresponding to 70% of dissimilarity at last level of fusion (0.94) (Figure 2). The first group is composed of the five sub-provenances (NMG, SMG, CMG, EMG and WMG) from MG; the second: PA, PE and PB while the third: SP and MS (Figure2).

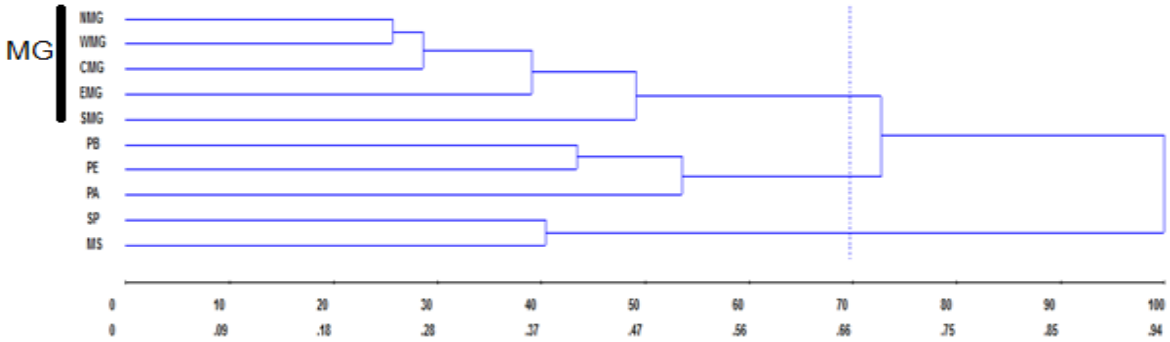

Figure 2. UPGMA dendogram of six *Acrocomia aculeata* provenances constructed from genetic distance (Nei, 1972) matrix. PA=Pará; PE=Pernambuco; PB=Paraiba; SP=São Paulo; MG (NMG=North Minas Gerais; SMG=south Minas Gerais; CMG=Central Minas Gerais; EMG=East Minas Gerais; WMG=West Minas Gerais); and MS=Mato Grosso do Sul. The first row (0-100) represents percentages of dissimilarity; and the second row (0-0.94) contains levels of fusion (average genetic distance) generated at each level of group formation.

However, with the optimization method of Tocher, one additional group was established forming the fourth group (or sub-group) due to the relative genetic dissimilarity of PA from its neighboring provenances (PE and PB) (Table 4). This is more likely because, unlike UPGMA, Tocher uses least genetic distances at each stage of group formation to establish new homogenous groups based on their genetic

similarities (Cruz *et al*., 2011). Hence, there is a possibility to encounter additional groups with method of Tocher.

Table 4. Grouping of *Acrocomia aculeata* provenances using method of Tocher

Group	Provenances	Mean intra-group distance	Mean inter-group distance
1	NMG WMG CMG EMG SMG	0.37	$D_{1,2}$=1.05; $D_{1,3}$=0.64; $D_{1,4}$=0.77
2	SP MS	0.38	$D_{2,3}$=0.80; $D2_{2,4}$=0.67
3	PB PE	0.41	$D_{3,4}$=0.50
4	PA	–	–

PA=Pará; PE=Pernambuco; PB=Paraiba; SP=São Paulo; MG (NMG=North Minas Gerais; SMG=south Minas Gerais; CMG=Central Minas Gerais; EMG=East Minas Gerais; WMG=West Minas Gerais); and MS=Mato Grosso do Sul. D is the mean inter-group distance.

According to Cruz *et al*. (2011), Tocher assumes establishment of optimum groups can be realized at 50% of dissimilarity. This could be confirmed from the dendogram (Figure 2) that, at less percentage of dissimilarity (50 to 55%), using a local criterion, PA formed as an independent and separate group reaffirming its relative genetic distance from the two neighboring provenances (PB and PE). Moreover, the relatively smaller genetic distance ($D_{3,4}$=0.50, Table 4) between group 3 (PB and PE) and group 4 (PA) demonstrated the genetic relatedness among the three provenances of the neighboring geographical States of North Brazil as also evidenced by the UPGMA grouping (Figure 2).

Results from the pair-wise genetic distance (or identity) matrix among the ten provenances (including five sub-provenances of MG) showed the maximum genetic distance (1.252) or minimum genetic identity (0.286) between South Minas Gerais (SMG) and São Paulo (SP); whereas, the minimum genetic distance (0.243) or maximum genetic identity (0.781) between North Minas Gerais (NMG) and West Minas Gerais (WMG) provenances (Table 5).

Table 5. Pair wise genetic distance (below diagonal) and genetic identity matrix (above diagonal) between *Acrocomia aculeata* provenances.

Provenances	PA	PE	PB	SP	NMG	SMG	CMG	EMG	WMG	MS
PA	-	0.676	0.603	0.445	0.469	0.481	0.528	0.515	0.553	0.415
PE	0.412	-	0.627	0.516	0.588	0.415	0.605	0.512	0.533	0.405
PB	0.547	0.467	-	0.46	0.415	0.543	0.401	0.445	0.492	0.563
SP	0.807	0.662	0.777	-	0.316	0.286	0.340	0.364	0.324	0.678
NMG	0.768	0.531	0.874	1.152	-	0.674	0.750	0.617	0.781	0.321
SMG	0.707	0.879	0.610	1.252	0.394	-	0.703	0.555	0.595	0.350
CMG	0.580	0.481	0.910	1.061	0.284	0.364	-	0.764	0.772	0.340
EMG	0.661	0.669	0.809	1.010	0.477	0.588	0.289	-	0.698	0.418
WMG	0.552	0.610	0.683	1.095	0.243	0.515	0.258	0.3436	-	0.386
MS	0.874	0.893	0.572	0.383	1.125	1.049	1.066	0.8695	0.886	-

PA=Pará; PE=Pernambuco; PB=Paraiba; SP=São Paulo; MG (NMG=North Minas Gerais; SMG=south Minas Gerais; CMG=Central Minas Gerais; EMG=East Minas Gerais; WMG=West Minas Gerais); and MS=Mato Grosso do Sul.

Comparison of genetic distance among families within each provenance demonstrated that the wider average genetic distance was among the families within Minas Gerais (0.668 = average of the five MG's sub-provenances) followed by within São Paulo (0.450) and Mato Grosso do Sul (0.280) (Table 6).

Table 6. Mean genetic distance among families within each provenance

Genetic distance*	SP	NMG	SMG	CMG	EMG	WMG	MS
Minimum	0.408	0.200	0.252	0.189	0.131	0.436	0.082
Maximum	0.522	1.410	0.961	1.718	0.901	1.276	0.568
Mean	**0.450**	**0.750**	**0.565**	**0.822**	**0.443**	**0.761**	**0.280**
SD	0.060	0.250	0.250	0.568	0.290	0.298	0.126

*These distances were not computed for PA, PE and PB as they have only single family.

The relatively narrow average genetic distance (0.280) observed in MS explained more genetic similarity among the families in this provenance compared to that in São Paulo (0.450) and Minas Gerais (0.668), respectively (Table 6). This

comparison was not done for Pará, Pernambuco and Paraiba since they have only one family.

Seven homogenous family groups were formed from the overall inter-family genetic distance analysis based on optimization method of Tocher (Table 7).

Table 7. Grouping of 41 *Acrocomia aculeata* families using Tocher from genetic distance (Nei, 1972).

Group	Families*	Provenance**
1	BGP118, BGP119, BGP103, BGP92, BGP105, BGP102, BGP117, BGP112, BGP104 BGP106, **BGP51, BGP34, BGP47**	MS/ **SP**
2	BGP11, BGP9, BGP52, BGP2, BGP21, BGP37, BGP78, BGP18, BGP33, BGP76 BGP24, BGP10	MG
3	BGP1, BGP3, BGP51, BGP49	MG
4	BGP20, BGP27, BGP64, **BGP82, BGP99, BGP124**	MG**/PE/PA/PB**
5	BGP14, BGP5, BGP68	MG
6	BGP22, BGP16	MG
7	BGP25	MG

*List of families and description of their origin can be inferred from Table 1.
MS= Mato Grosso do Sul; SP=São Paulo; MG=Minas Gerais; PE=Pernambuco; PA=Pará; PB=Paraiba. Families in **bold are those which formed groups with other families in group 1 and 4.

The first group is composed of all families (100%) from Mato Grosso do Sul and São Paulo. This reaffirmed the close relatedness among families from the latter two provenances as already explained in figure 2 (UPGMA), figure 3 (PCoA), figure 4 and table 4 (Tocher) using different grouping methods. This homogenous grouping of families from Mato Grosso so Sul and São Paulo provenances was also explained by the smallest average genetic distance obtained among the families within these two provenances, respectively (Table 6).

Group 2, 3, 5, 6 and 7 are different homogenous groups formed by families from Minas Gerais demonstrating the diverse family groups in Minas Gerais collections (Table 7). Group 4 is composed of a mixture of families from Minas Gerais, Pará, Pernambuco and Paraiba, confirming the genetic relatedness beween the latter four

provenances as already explained by grouping among provenances using UPGMA (Figure 2, at > 75% of dissimilarity; and 3D graphical dispersion using principal components, Figure 4).

Unlike UPGMA and Tocher, the graphical method of grouping- principal coordinate analysis (PCoA) did not clearly distinguish the three groups, which were clearly separated by the other two methods (Figure 3).

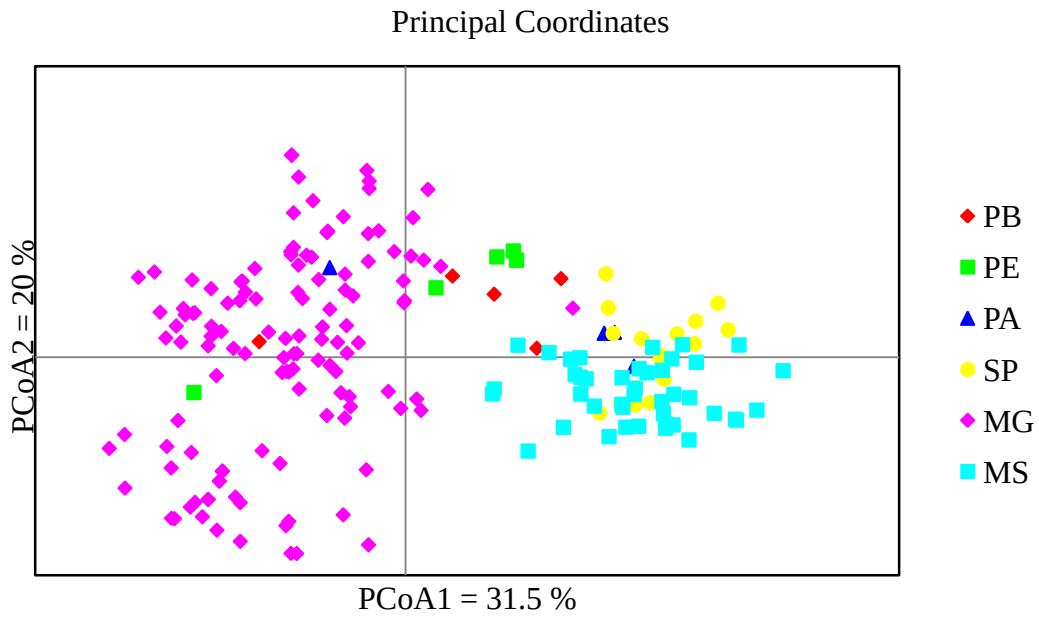

Figure 3. Graphical dispersion of 192 *Acrocomia aculeata* accessions using Principal Coordinate Analysis (PCoA) showing grouping of the accessions into different distinct groups. Provenances include PA=Pará, PE=Pernambuco; PB=Paraiba; SP=São Paulo; MG=Minas Gerais (containing five sub-provenances: NMG, SMG, CMG, EMG and WMG shown as on group); and MS= Mato Grosso do Sul.

However, PCoA completely separated Minas Gerais collections from that of São Paulo and Mato Grosso do Sul, in which the latter two provenances formed one separate group, which is consistent in the other two methods (Figure 2 and Table 4). Nevertheless, since PCoA explained only 51.5% of the total variability on the bi-dimensional axes (PCoA1=31.5%; PCoA2=20%) (Figure 3), this method was less efficient be used in grouping *A. aculeata* accessions in this study.

However, graphical dispersion of the provenances on principal components on 3D projection showed high correlation (r= 0.75), and reduced distortion (4.98%) and stress (29.36%), which make this method more efficient to explain the groupings of the provenances (Figure 4). This method established the three groups and was consistent with the other two methods (UPGMA-Figure 2; Tocher-Table 4).

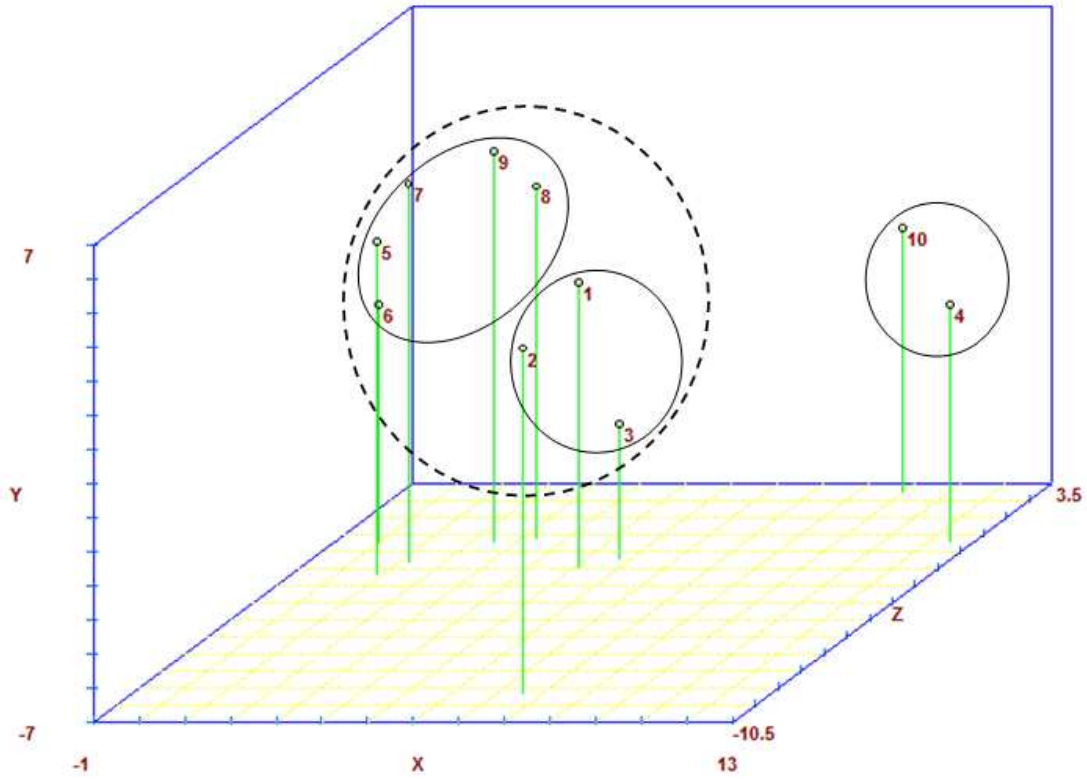

Figure 4. Graphical dispersion of provenances of *Acrocomia aculeata* in relation to three principal components, established by analysis of 10 SSR loci. Provenances include 1=Pará, 2=Pernambuco; 3=Paraiba; 4=São Paulo; 5−9 are sub-provenances of Minas Gerais (5=NMG; 6=SMG; 7=CMG; 8=EMG; 9=WMG) and 10=Mato Grosso do Sul.

Establishment of the different groups reiterates the presence of diverse genetic variability among the accessions in the germplasm bank. Besides, the consistency of the groupings using three different methods demonstrated, the groups established are genetically distinct and could be considered in the pre-breeding program of *A. aculeata* in the germplasm bank

Our results of the grouping detected genetic similarities among *A. aculeata* accessions among some geographically distant regional States in Brazil; for instance between MG and that of PB, PE and PA (Figure 2, at > 75% of dissimilarity, Figure 4). On the other hand, complete genetic dissimilarity was detected between accessions from MG and that of the neighboring SP and MS (Figure 2, Figure 4). This scenario was also explained by the mean inter-group genetic distances with Tocher (Table 4). The highest average genetic distance ($D_{1,2}$=1.05) was obtained between group 1 (MG) and group 2 (MS, SP) than between group 1 and group 3 (PE, PB; $D_{1,3}$=0.64) and group 1 and 4 (PA; $D_{1,4}$=0.77). This result was reaffirmed by a Mantel correlation test detecting a non-significant correlation (r=0.4128) between genetic and geographic distances among the provenances studied (Figure 5).

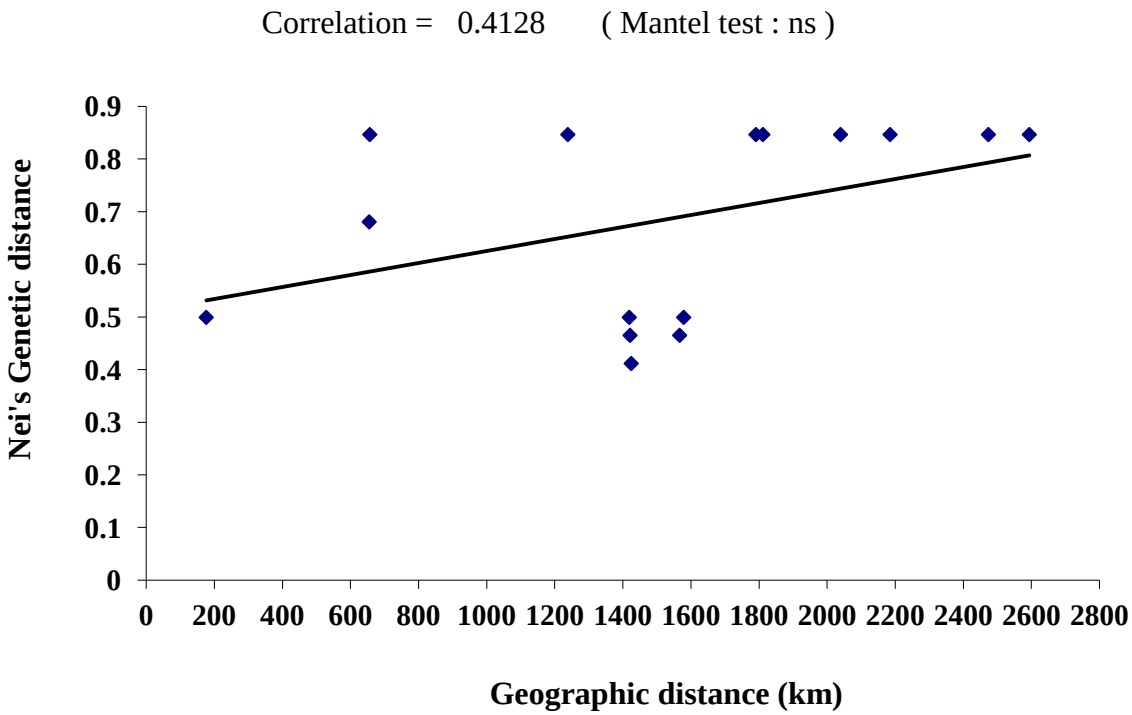

Figure 5. Mantel test revealed correlation between genetic and geographic distance among provenances.

Increasing geographic distances does not necessarily increase genetic distances and vice versa (Figure 5). The non-significant correlation explained the results of

genetic similarities among some geographically distant provenances analyzed in the study and genetic dissimilarities among nearby provenances. For example: genetic similarities among Minas Gerais, Pará, Pernambuco and Paraiba and genetic dissimilarities among Minas Gearis, São Paulo and Mato Grosso do Sul.

The genetic similarities and differences among the provenances signified the dispersal of *A. aculeata* in Brazil was prominent. Although we were not intended to study the dispersal of *A. aculeata* in Brazil, we tried to relate our results with some archeological findings in the past. According to Silva and Russo (2000), there are evolutionary forces that might cause the differentiation of populations within time and geographical distribution of species, such as natural selection, genetic drift, migration and phylogenetic and biogeographic patterns of the specie.

However, archeological remains suggested that *A. aculeata* is one of the five best-represented palms at several archaeological sites in South-America including Brazil. Remains of endocarps and seeds revealed fruits were the main agents for its dispersal by humans (Morcote-Rios and Bernal, 2001). There are even good reasons to think that the palm was dispersed during that time by humans: First, the fruits of *A. aculeata,* unlike those of other oil-producing palms, have an abundant mesocarp that can be consumed directly, without a long and time consuming oil-extraction process. Second, the thick, brittle exocarp offers good protection for the mesocarp and can be easily peeled off when necessary. Third, the fruits do not ferment quickly, thus allowing their consumption for a period of several weeks. Thus, it would seem a good choice for migrating people to take fresh fruits of *Acrocomia* for eating first the fleshy and oily mesocarp during travel and then the seed (Levi-Strauss, 1950; Janzen, 1983).

Besides, it was discovered that there has been a North-ward dispersal of the *A. aculeata* from South-America to Central America using the age of the remains studied at different archeological sites (Morcote-Rios and Bernal, 2001). In our results of the

grouping using dendogram (Figure 2), at higher percentage of dissimilarity (> 75%), the genetic relatedness between MG's accessions and that of the Northern regional States including PA, PB and PE could implicate the North-ward dispersal of the *A. aculeata* in time and space. Similarly, the genetic similarities between the neighboring States of SP and MS (Figure 2; Figure 4; Table 4) could be accounted for the dispersal of genetic materials (mainly fruits) by humans or by the help of rivers that flow between the two States. However, the complete dissimilarities between genetic materials of MG and that of the neighboring SP and MS could raise a question why they are totally dissimilar. We suggest that, there might be possibilities to encounter other natural factors such as long chains of mountains or absence of rivers flowing between these regions that might hinder the migration of the genetic materials between them. Therefore, this would lead to a recommendation that further evolutionary studies such as phylogeography is required to reach at a definite conclusion on the dispersal of *A. aculeata* within Brazil.

Three hierarchal levels of analysis of molecular variance (AMOVA) detected a higher genetic variation within family (48.22%) than among families (36.47%) in accessions of *A. aculeata* followed by variation among provenances (15.31%) respectively (Table 8).

These variations were statistically significant ($p < 0.01$) in that all Φ-statistics' values obtained from the estimated variances were higher than those values obtained under 1000 permutations (not in the table). The results confirmed that within family, genetic variation is accounted for the highest share as compared to among families (with in a provenance) variation followed by among provenance variation. The smallest share of genetic variation among the provenances revealed the genetic similarities obtained among the provenances such as between (MS and SP) and (MG and PB, PE and PA) (Figure 2; Table 4).

Some previous studies in *A. aculeata* reported higher genetic variability within population than among populations (Oliveira *et al*., 2012; Colombo *et al*., 2013).

Table 8. Analysis of molecular variance (AMOVA) performed for 178 *Acrocomia aculeata* accessions based on ten polymorphic SSR loci.

Source of variance	df	Components of Variance	Percentage of variance	**Φ**-Statistics	Sig.
Among provenances	6	0.1217	15.31	Φ_{GT}= 0.1531	*
Among families	31	0.2901	36.47	Φ_{SG}= 0.4307	*
Within family	140	0.3835	48.22	Φ_{ST} = 0.5178	*
Total	177	0.7953	100		

*significant (P < 0.01) with null hypothesis: H_0: σ_a^2=σ_b^2=σ_c^2=0; Φ-statistics are compared with values obtained from 1000 permutations. AMOVA performed using 178 accessions of 38 families from seven provenances (regions) including SP=São Paulo; MG (NMG=North Minas Gerais; SMG=south Minas Gerais; CMG=Central Minas Gerais; EMG=East Minas Gerais; WMG=West Minas Gerais) and MS=Mato Grosso do Sul. Parã, Pernambuco and Paraiba were not included in the analysis since they have only one family.

According to Colombo *et al*. (2013), the higher genetic diversity within population of *A. aculeata* is a result of the mixed mating system in *A. aculeata* and the involvement of metapopulation structure in natural population. Metapopulation structure of specie could be caused by fragmentation of lands and creates spatially separated populations which interact at some level. However, this could favor genetic drift and restricted gene flow, which cause decrease in genetic diversity within a population especially in cross pollinated species (Levins, 1969). However, in our case, the mixed mating nature of *A. aculeata* (Abreu *et al*., 2012) would keep the genetic variability higher within a family regardless of the influence of metapopulation structure in natural populations. Similar works also reported higher genetic variation within populations in some palm and different tree species: *Phoenix canariensis* (González-Pérez *et al*., 2004), *Populus tremuloides Michx* (Yeh *et al*., 1995), *Digitalis minor*

(Sales *et al*., 2001), *Piper hispidinervum* (Wadt, 2001) and *Trichilia pallida* (Zimback *et al*., 2004).

Since this study was based on active collections of accessions from various provenances representing vast geographical areas in the country, there is high chance of encountering more than one species of genus *Acrocomia*, which still have not got specific descriptors to distinguish them. However, the result confirmed the presence of high genetic variability in the collections which could be at least taken as preliminary information to plan appropriate conservation and management strategies of genetic resources in the collections, which will in turn contribute to the breeding program. Conducting more additional studies with strategically designed collections; more population size and with more number of SSR markers would provide information that would help the breeding program of *A. aculeata*.

4. CONCLUSIONS

The following conclusions could be drawn from the results obtained:

The higher genetic diversity observed among the families within a provenance attested that collections from a provenance are more advantageous than collections from all provenances.

Reduced average observed heterozygosity per locus (or provenance) has implications that populations of mother plants, which gave rise to the accessions studied here, were suffered from genetic drift due to antropization or restricted gene flow.

Establishment of three groups of *A. aculeata* accessions which were confirmed by UPGMA, Tocher and 3D graphical disperssion confirmed the presence of genetic variability in the germplasm bank collections and this could be useful in early stages of pre-breeding to form basic populations.

Genetic relatedness (or similarity) among some geographically distant provenances; for example between Minas Gerais, Pernambuco, Paraiba and Pará; demonstrating that *A. aculeata* has been dispersed for long distances within Brazil.

These data are preliminary and further studies should be performed, covering larger number of individuals, populations and more number of markers.

5. REFERENCES

Abreu SI, Carvalho CR, Carvalho GMA and Motoike SY (2011) First karyotype, DNA C-value and AT/GC base composition of macaw palm (*Acrocomia aculeata*, Arecaceae) –a promising plant for biodiesel production. *Australian Journal of Botany* 59: 149–155.

Abreu AG, Priollig RH, Azevedo-Filho JA., Nucci SM, Zucchi SM, Coelho RM and Colombo CA (2012) The genetic structure and mating system of *Acrocomia aculeata* (Arecaceae). *Genetics and Molecular Biology* 35: 119–121.

Bora PS and Rocha RVM (2004) Macaíba palm: fatty and amino acids composition of fruits. *Ciência e Tecnologia de Alimentos* 4: 158–162.

Bostein D, Whiter L, Skolnick M and Davis RW (1980) Construction of a Genetic Linkage Map in Man using Restriction Fragment Length Polymorphisms. *The American Journal of Human Genetics* 32: 314–331.

Brito GG, Caixeta ET, Gallina PA, Zambolim EM, Zambolim L, Diola V and Loureiro ME (2010) Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* 173: 255–264.

Cole CT (2003) Genetic variation in rare and common plants. *Annual Review of Ecology, Evolution and Systematics* 34: 213−237.

Colombo CA, Nucci SM, Priolli RHG, Zucchi MI, Carhalho CRL, Chorfl BLH, Siqueira WG and Azvedo-Filho JÁ (2013) Genetic Diversity of Macaw Palm (*Acrocomia aculeata*) by Microsatellite Markers. International Plant and Animal Genome Conference. Available at http://www.intlpag.org.

Cruz CD (2013) GENES V9.0. *Software package for analysis in quantitative genetics and experimental statistics*, Universidade Federal de Viços, Brazil.

Cruz CD, Ferreira FM and Pessoni LA (2011) Biometria aplicada ao estudo da diversidade genética.Viçosa: Editora Universidade Federal de Viçosa, p.622.

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

Ersts PJ. Geographic Distance Matrix Generator (v.1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available at http://biodiversityinformatics.amnh.org/open_source/gdmg. Accessed on 23/12/2014.

Excoffier L (2001) Analysis of population subdivision, IN: Balding DJ, Bishop M and Cannings C(Eds.) *Handbook of statistical genetics*. Chichester (UK), pp. 271-307.

Excoffier L, Smouse PE and Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

Faleiro FG, Costa AM, Karia CT, Andrade RP, Junqueira NTV, Pereira AV, Pereira EB and Sano SM (2008) Molecular markers and geographic information systems as a tool to study native plant species in the Brazilian Savannas. In: *The 2nd International Symposium in Tropical Savannas*. 12−17 October 2008, ParlaMundi, Brazilia. Available at http://www.cpac.embrapa.br.

Ferreira ME (2005) Molecular Analysis of Gene banks for sustainable conservation and increased use of crop genetic resources. *In*: *The role of biotechnology*. Villa Gualino, Italy. Available at http://www.fao.org/biotech.

Fortes ICP and Baugh PJ (1999) Study of analytical on-line pyrolysis of oils from Macauba fruit (*Acrocomia sclerocarpa* M.) via GC/MS. *Journal of The Brazilian Chemical Society* 10: 469–477.

González-Pérez MA, Caujapé-Castells J and Sosa PA (2004) Allozyme variation and structure of the Canarian endemic palm tree *Phoenix canariensis* (Arecaceae): implications for conservation. *US National Library of Medicine National Institute of Health.*

Hiane PA, Ramos-Filho MM, Ramos MIL and Macedo MLR (2005) Bocaiúva, *Acrocomia aculeata* (Jacq.) Lodd., pulp and kernel oils: characterization and fatty acid composition. *Brazilian Journal of Food Technology* 8: 256–259.

Hartl DL and Clark AG (1997) *Principles of population genetics*. USA: Sinauer Associates, p.542.

Henderson A, Galeano G and Bernal R (1995) Field guide to palms of the Americas, New Jersey: Princepton University, pp. 166-167.

Hoisington D, Khairallah M and Gonzalez-de-Leon D (1994) Laboratory Protocols: CIMMYT Applied Biotechnology Center. Second Edition, Mexico, D.F.: CIMMYT.

Janzen D (1983) *Acrocomia vinifera* (Coyol). In: Janzen DH (ed.) *Costa Rican natural history.* Chicago, pp. 184−185.

Lanes ECM, Nick C, Kuki KN, Freitas RD and Motoike SY (2013) Genomic DNA isolation of *Acrocomia aculeata* (Arecaceae) from leaf and stipe tissue samples for PCR analysis. *Genetics and Molecular Research* 12: 3905-3911.

Lanes ECM, Costa PMA and Motoike SY (2014) Alternative fuels: Brazil promotes aviation biofuels. *Nature* 511:31.

Levi-Strauss C (1950) The use of wild plants in tropical South America. In: Steward J H (ed.) *Handbook of South American Indians.* Washington, pp. 465−486.

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15*:* 237–240.

Mengistu FG, Motoike SY, Caixeta ET, Cruz CD and Kuki KN (2015) Cross-species amplification and characterization of new microsatellite markers for the macaw palm, *Acrocomia aculeata* (Arecaceae). Plant Genetic Resources. doi: doi:10.1017/S1479262115000179.

Motoike SY, Lopes FA, Sá Júnior AQ, Carvalho M and Oliveira MAR (2007) Processo de Germinação e Produção de Sementes Pré-Germinadas de Palmeiras do Gênero *Acrocomia*. Submetido à Lei de Patentes. Protocolo INPI: 014070005335.

Motoike SY and Kuki KN (2009) The potential of macaw palm (*Acrocomia aculeata*) as source of biodiesel in Brazil. *International Review of Chemical Engineering* 1: 632–635.

Morcote-Rios G and Bernal R (2001) Remains of Palms (Palmae) at Archaeological Sites in the New World: A Review. *The Botanical Review* 67: 309-350.

Moura EF, Motoike SY, Ventrella MC, Sá Junior AQ and Carvalho M (2009) Somatic embryogenesis in macaw palm (*Acrocomia aculeata*) from zygotic embryos. *Scientia Horticulturae* 119: 447–454.

Nei M (1972) Genetic distance between populations. *American Naturalist* 106:238-292.

Nei M (1978) Estimation of average heterozygosity and genetic distances from a small number of individuals. *Genetics* 89: 583–590.

Nucci SM (2007) Development, characterization and analyses of the use of microsatellite markers in macaw palm population genetics. MSc. Thesis, University of Sâo Paulo.

Nucci SM, Azevedo-Filho A, Colomo AC, Priolli GHR, Coelho MR, Mata TL and Zucchi IM (2008) Development and characterization of microsatellites markers from the macaw. *Molecular Ecology Resources* 8: 224–226.

Oliveira DA, Júnior-Melo AF, Brandão MM, Rodrigues LA, Menezes EV and Ferreira PRB (2012) Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais. *Genetics and Molecular Research* 11: 531–538.

Peakall R and Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Ramos SF, Vasconcelos de Macêdo JL, Lopes MTG, Batista JS, Formiga KM, Pimentel da Silva P, Saulo-Machado AC and Veassey EA (2012) Microsatellite Loci for Tucumã of Amazonas (*Astrocaryum aculeatum*) and amplification in other Arecaceae. *American Journal of Botany* 99: e508-e510.

Read RW (1966) New chromosome counts in the Palmae. *Principles* 10: 55-61.

Ribeiro LM, Souza PP, Rodrigues AG, Oliveira TGS and Garcia QS (2011) Overcoming dormancy in macaw palm diaspores, a tropical species with potential for use as bio-fuel. *Seed Science and Technology* 39: 303–317.

Sales E, Nebauer SG, Mus M and Segura J (2001) Population genetic study in the Balearic endemic plant species *Digitalis minor* (Scrophulariaceae) using RAPD markers. *American Journal of Botany* 88: 1750−1759.

Scariot AO, Lleras E and Hay JD (1991) Reproductive biology of the palm *Acrocomia aculeata* in Central Brazil. *Biotropica* 23: 12–22.

Scariot AO, Lleras E and Hay JD (1995) Flowering and fruiting phenologies of the palm *Acrocomia aculeata*: patterns and consequences. *Biotropica* 27: 168-173.

Silva EP and Russo CAM (2000) Techniques ans statistical data analysis in molecular population genetics. *Hydrobiologia* 420: 119-135.

Tanksley YD, Young ND, Paterson AH and Bonierbal MW(1989) RFLP mapping in plant breeding: new tools for an old science. *BioTechnology* 7:257–264.

Tickel J (2000) From the Fryer to the Fuel Tank: The Complete Guide to using Vegetable Oil as an Alternative Fuel. Florida: Tickel Energy Consulting, p.155.

Wadt LH (2001) Estrutura genética de populações naturais de Pimenta longa (*Piper hispidinervum*), visando seu uso e coservação. Ph.D. Thesis, Universidade de Federal de São Paulo.

Yeh FC, Chong DK and Yang RC (1995) RAPD variation within and among natural populations of trembling aspen (*Populus tremuloides* Michx.). US National Library of Medicine National Institute of Health, Alberta.

Zimback L, Mori ES, Kagyama PY, Veiga RFA, *et al*. (2004) Estrutura genética de populações de *Trichilia pallida* Swartz (Meliaceae) por marcadores RAPD. *Scientia Forestalis* 65: 114-119.

GENERAL CONCLUSIONS

Transferability of SSR primers between species in Arecaceae could be used as an alternative means to select working primers for *A. aculeata*.

The primers selected from the transferability could be applied to realize various studies in *A. aculeata*, such as population genetics, germplasm characterization, genetic improvement and conservation.

The results indicated that there are possibilities to increase the success of transferability if sourcing SSR primers at genus or species level.

The study also confirmed that SSR are very useful and efficient molecular markers to characterize *A. aculeata* accessions in the germplasm bank.

The information generated in the results would facilitate the process of identifying, grouping and selecting genetically divergent groups during the course of pre-breeding in the germplasm bank.

The differences in amount of genetic variability obtained among provenances and among families within a provenance would help to plan appropriate collection and conservation strategies and optimize the use of the genetic resources in the germplasm bank for future breeding programs.