

## Genetic selection of *Calophyllum brasiliense* for seed orchards

Guilherme Schnell Schühli<sup>1\*</sup>, Thiago Wendling Gonçalves de Oliveira<sup>2</sup>, Maria do Socorro Padilha de Oliveira<sup>3</sup>, João Antonio Pereira Fowler<sup>1</sup>

### ABSTRACT

*Calophyllum brasiliense* populations are under severe depletion and criteria to improve production and quality of propagative material are therefore necessary. Genetic measures have the potential to reduce consanguinity and maximize allelic representation within target populations. Here, we explored genetic values for this species in a small relic of natural forest in Rio de Janeiro State (Brazil). The objective was to evaluate the potential of some genetic measures for seed orchard establishment. As genomic information of native trees is still scarce, we opted to use a dominant marker: RAPD. DNA from 17 phenotypically superior trees was obtained through the CTAB method and submitted for amplification by PCR. Electrophoresis and electronic documentation was then conducted. We calculated the percentage of polymorphic bands (PPB), gene diversity ( $H_t$ ), Shannon's information index ( $i$ ), genetic distance (UPGMA) and parsimony analysis. Six primers were evaluated generating 34 loci. We found high genetic diversity PPB=70.6% with  $H_t=0.28$  and  $i=0.41$ . Genetic relationships were reported in dendrograms (maximum parsimony and distance). Simulated sampling within and among clusters suggests that inter cluster sampling is more effective to capture the genetic diversity.

**Key-words:** RAPD, propagative material, MAS, guanandi.

## Seleção genética de *Calophyllum brasiliense* para a formação de pomares de sementes

### RESUMO

As populações de *Calophyllum brasiliense* encontram-se sob severa depleção e, portanto, são necessários critérios para melhorar a produção e a qualidade do material propagativo da espécie. Índices genéticos têm o potencial para orientar a redução de consangüinidade e de maximizar a representação alélica dentro das populações de interesse. Neste artigo exploramos valores genéticos para esta espécie em um pequeno relicto de floresta natural no Estado do Rio de Janeiro (Brasil). O objetivo foi o de avaliar o potencial de algumas medidas genéticas para o estabelecimento de pomares de sementes. Desde que informações genômicas de árvores nativas ainda são escassas optamos pelo uso de um marcador dominante: RAPD. O DNA de 17 árvores fenotipicamente superiores foi obtido através do método CTAB e encaminhado para amplificação por PCR. Eletroforese e documentação eletrônica foram então conduzidas. Calculamos a porcentagem de bandas polimórficas (PPB), diversidade genética ( $H_t$ ), índice de informação de Shannon ( $i$ ), distância genética (UPGMA) e análise de parcimônia. Seis iniciadores foram avaliados gerando 34 loci. Encontramos alta diversidade genética PPB=70,6% com  $H_t=0,28$  e  $i=0,41$ . As relações genéticas foram apresentadas em dendrogramas (Máxima parcimônia e distância). Amostragem simulada dentro e entre agrupamentos sugerem que a amostragem dentro de grupamentos é mais eficiente para melhor capturar a diversidade genética.

**Palavras-chave:** RAPD, material de propagação, MAS, guanandi

\*Author for correspondence.

<sup>1\*</sup>Embrapa Forestry - Estrada da Ribeira, km 111, Po-Box 319 - Colombo, PR - Brazil - 83411-000 Phone: 55 (41) 3675-5789 - Fax: 55 (41) 3675-5601. guilherme.schuhli@embrapa.br \*, joao-antonio.fowler@embrapa.br

<sup>2</sup>Federal University of Paraná grad. student. thiagowendling@yahoo.com.br

<sup>3</sup>Embrapa Eastern Amazon - Trav. Dr. Enéas Pinheiro s/nº Po-Box, 48 Belém, PA - Brazil CEP 66095-100 spadilha@cpatu.embrapa.br

## INTRODUCTION

*Calophyllum brasiliense* Cambess. (Clusiaceae) is found in many of Brazil's most representative biomes, including the Amazon, Atlantic forest and Cerrado (Marques and Joly 2000). Its geographical distribution spans from Central America to southern coast of Brazil (Santa Catarina State). *C. brasiliense* is adapted to high soil humidity and is typically found in waterlogged forests (Reitz et al. 1978; Oliveira-Filho and Ratter 1995). Natural populations of *C. brasiliense* (popularly known as guanandi) are under severe depletion due to illegal harvesting and deforestation driven by the rapid expansion of the agricultural frontier (Marques and Joly 2000).

Recent reformulations of the Brazilian laws controlling agricultural activities and natural conservation areas have led to renewed interest in the restoration of natural forests. Indeed, rehabilitation of forests is a requirement of the new Brazilian Forestry Code. Effective restoration heavily depends on the wide availability of propagative material for a range of native species. However, seedlings and even seeds of native tree species are commercially scarce in Brazil (Caldas 2006). Moreover, the seedlings and seeds that are available are not necessarily of the highest quality (e.g. phytosanitary and genetic quality).

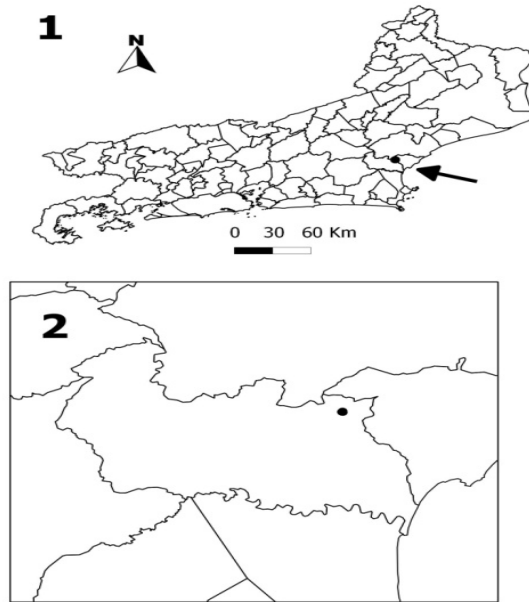
Improving the production of propagative material is especially important for the effective rehabilitation of native forests. Genetic measures are an especially effective strategy to increase production (Sebbenn 2006). For example, genetic information can be used to direct sampling efforts to reduce inbreeding and the loss of diversity through genetic drift, thereby improving the quality of the future seed orchard composition. Thus, with a basic knowledge of genetic parameters, trees are selected for propagation not only based on their productivity and health characteristics, but also considering consanguinity and broad allelic representation within target populations (Sebbenn 2002).

The objective of this study was to assess the potential genetic measures to improve the production of *Calophyllum brasiliense*. A small fragment of natural forest in Rio de Janeiro State (Brazil) was selected as source of matrix trees to provide seeds for propagation. After phenotypic selection of 'superior' trees, we calculated genetic diversity parameters to facilitate the development of optimal sampling strategies for the creation of seed orchards. This is the first population study of this species based on DNA markers and provides important baseline information for future research to assure genetic quality in seed and seedling production of guanandi.

## MATERIAL AND METHODS

Genetic assessment was through a simple molecular marker. As genomic information of native Brazilian tree species is still scarce and there is the need for a fast processing time, we used dominant markers. Moreover, to ensure the methods adopted could, if necessary, be scaled-up, it was also important to use a transferable marker with low costs and with a simple procedure. Random amplification of polymorphic DNA – RAPD fulfills these characteristics, providing estimation of minimal population parameters with high transferability within native species. RAPD is also fast, technically simple and allows for a high frequency of identification of polymorphisms (Newbury and Ford-Lloyd 1993). However, a possible disadvantage of RAPD is low reproducibility when researchers lack tight experimental control of the PCR conditions (Muchugi et al. 2009).

Material for determining genetic variability was collected from 17 phenotypically 'superior' trees (=apparently healthy and with a high diameter at breast height) located in a small conservation area in *Casemiro de Abreu* municipality in the northeast of the Brazilian State of Rio de Janeiro (Figure 1).



**Figure 1.** Location of Rio de Janeiro State and Casimiro de Abreu municipality (1) and the experimental area (2).

Young and healthy leaves from each individual were collected and were stored in plastic bags with silica gel for transportation to the laboratory. Leaf samples were stored in a freezer at  $-20^{\circ}\text{C}$ .

Genomic DNA was obtained from cold stored leaves using a CTAB (cetyltrimethylammonium bromide) based protocol (Cheung et al. 1993). After extraction, the genomic extract was eluted in bi-distilled water and stored at  $-80^{\circ}\text{C}$ . DNA samples were submitted to the polymerase chain reaction (PCR)-RAPD and exsiccate and genomic DNA were deposited in Embrapa's Florestas Herbarium and Molecular Biology Collection as vouchers.

Amplifications were replicated to ensure reproducibility. All samples were submitted to 3 PCRs (at least) for each primer set. Only bands reproducible in two independent amplification reactions were included in the data analyses. The amplification reactions were carried out in a final volume of 25  $\mu\text{L}$ . PCR reactions consisted of 1x PCR buffer 2mM  $\text{MgCl}_2$ , 0.2mM dNTPs (each),  $\mu\text{M}$  0.5 desired oligonucleotide (primer), 1U Taq polymerase, and water to complete 25  $\mu\text{L}$ . Template DNA was adjusted to 2 ng/ $\mu\text{L}$  in the PCR reaction tube. We tested 10 oligonucleotides (Operon Technologies Inc.). The reactions were amplified in an Applied

Biosystems thermal cycler model Veriti 96 Well. Amplifications were performed on a thermal cycler with 4 minutes of initial denaturation at  $92^{\circ}\text{C}$ ; 40-cycles consisting of 1 minute at  $92^{\circ}\text{C}$ , 1m and 30s at  $40^{\circ}\text{C}$ , and 2m at  $72^{\circ}\text{C}$ ; and final extension of 5m at  $72^{\circ}\text{C}$  for the complete extension of amplified products. Default electrophoresis procedures (1,5% agarose; ethidium bromide staining; 80 V) and electronic documentation was conducted to register PCR results.

RAPD data are dominant, and each band therefore represented the phenotype at a single bi-allelic locus. To minimize biased interpretation of the electrophoresis gels, we implemented automatic gel reading with a Gel Analyzer v. 2010a (Lazar 2010). After background subtraction (manual baseline tool - Gel Analyzer), the multiple gels from same primer set (replicates) were visually compared to confirm band patterns. A binary matrix was then composed based on this evaluation.

We examined sampling variance to determine how many markers would be required for a given level of precision in the estimate of genetic distances (Tivang et al.1994; Manly 1997). The procedure was done within the software GENES v. 2009.7.0 (Cruz 2006), considering 1 to 32 markers and similarity under Nei and Li (1979). The ideal number of polymorphic bands was based on the suggested  $< 0.05$  stress value (Kruskal 1964). The percentage of polymorphic bands, Nei's gene diversity, Shannon's information index, and genetic distance were calculated using POPGENE 1.32 (Yeh et al., 1997) and were based on the Hardy-Weinberg equilibrium.

Pairwise genetic distances were calculated using the Nei coefficient (Nei, 1972). Phenograms were prepared based on UPGMA (unweighted pair group method using arithmetic averages) (algorithm FIND activated with maximum number of tied trees set to 25). The genetic distance matrix was also subjected to a principal coordinate analysis (PCA). New independent axial coordinates were calculated from the genetic distances, which represent most of the variability of the original data. The taxa were then plotted as points in three dimensional continuous space defined by the first three coordinates. These calculations were done using the NTSYS-PC program (Rohlf 1993). Neighbour-joining (NJ) search was computed

with PAUP\*4b10 (Swofford 2003). As distance methods sometimes perform poorly to identify evolutionary patterns, we also included a heuristic search for an optimal tree. Maximum parsimony analyses (MP) of the binary data matrix were performed with PAUP\* v. 4b10 using the heuristic search (MULPARS, ACCTRAN, random sequence addition, 10,000 replicates, and TBR branch swapping). Negative branch lengths were allowed, but set to zero for tree-score calculation. Steepest descent options were not in effect. Starting tree(s) were obtained via neighbour joining. Data was treated as unrooted. Bootstrap analysis (Felsenstein 1985) was also performed to assess statistical support for branches in the most parsimonious trees and neighbour-joining trees. Branch support was evaluated for 10,000 replicates with neighbour joining. Bremer support (Bremer 1994) was calculated for parsimony trees.

Based on the results of the parsimony and distance analysis, we selected individual trees based on their genetic relationships. Genetic diversity indexes within groups of closely related individuals were compared with indexes obtained from groups of distantly related individuals. Simulated diversity indexes obtained from both sampling strategies were considered based in Nei's diversity index.

## RESULTS AND DISCUSSION

Six primers resulted in successful amplification. Selected oligonucleotide sequences (provided below) may be useful for further investigation within the genus (Table 1). Unsuccessful PCR resulted from the primers OPN02, OPN03, OPP19, and OPM02.

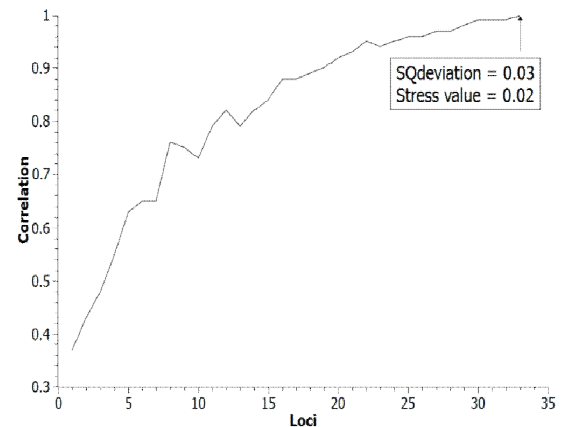
**Table 1.** Oligonucleotide used in PCR, respective sequences and loci number.

Oligonucleotide	Sequence (5'-3')	Loci number
OPA01	CAGGCCCTTC	8
OPC10	TGTCTGGGTG	7
OPE20	AACGGTGACC	3
OPJ19	GGACACCACT	5
OPP13	GGAGTGCCTC	11

Phenotypic data was organized in a binary matrix with 34 loci for 17 individuals. We found 24 (70.6%) polymorphic loci with an average of

1.7 observed and 1.5 effective alleles per loci. Nei's gene diversity was 0.28 and Shannon's Information index was 0.41 indicating that the assessed *C. brasiliense* population displayed a high diversity. OPT13 amplified the largest number of loci (11) while OPE20 amplified the least (3).

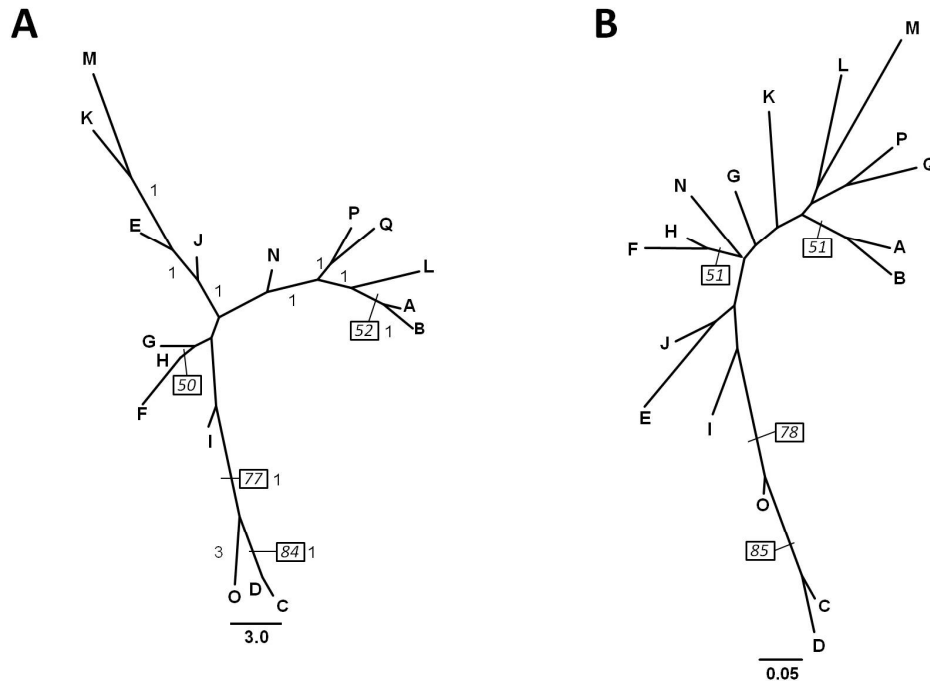
Estimations of precision with 34 loci revealed high correlation values, reaching 0.8 with 14 loci, 0.9 with 20 and 1 with 33. The stress value with 34 loci was less than 0.05 (Figure ).



**Figure 2.** Average bootstrapped correlation values and evaluated loci number. Square deviation and stress values for the 33 loci were 0.03 and 0.02, respectively.

Assuming the Hardy-Weinberg equilibrium, the expected diversity at species level was 0.2030. Shannon's index of diversity was 0.4144. A simple heuristic search under by parsimony yielded two unrooted topologies (consistency index 0.35; homoplasy index 0.64; and retention index 0.54). The conflict was restricted to the arrangement of the samples F, G and H.

We opted for one of the 2 topologic resolutions instead of the strict consensus to keep branch length values (Figure A). The groups are topologically congruent with the definition found in classical UPGMA analysis (at 20% of genetic distances). Group E, K, and M and F, G, and H (first structured under parsimony analysis) were grouped in UPGMA with A, B, L, M, P and Q. Bootstrap support is shown in Figure (A and B) and Bremer support values are shown in Figure (A).



**Figure 3.** Genetic relationships as represented by unrooted dendrograms. A Parsimony topology (IC=0.35; HI=0.64; RI=0.54), branch length represent character changes; B Distance topology (UPGMA), branch length represents distances. Bootstrap values (10,000 replicates), when > 50% are presented in boxes, Bremer support (for parsimony only), when more than 0 were near respective branch.

The low consistency and retention index indicate considerable misidentification of loci as homologous. The lower bootstrap values (< 50%) for some branches, under both searches strategies (UPGMA and Parsimony), demonstrate incongruence derived from non-homology loci. The unsolved position for these branches (including alternatives topologies for F,G and H) are directly related to the low consistency and retention index. Nonetheless, branches with higher bootstrap values (and confirmed in both analyses) are robust and not affected by non-homologous loci data.

Even taking into consideration our limited geographical sampling area, both genetic diversity measures were high (Nei's and Shannon's diversity index). This suggests that dominant markers are still able to recover recent

genetic relationships specifically for undomesticated tree species. These species are generally reported as having high genetic diversity, which can be observed in the easy amplification of primers and in the polymorphic loci. It should be noted that the unsuccessful primers were immediately discarded to allow fast processing. We believe that a careful optimization could raise the viability of some of these discarded markers.

A reasonable precision for this scale of evaluation was obtained with the 34 obtained loci. The correlation and stress values support the observed topologies. Although it is possible to examine genetic differences based on RAPD data, support values suggest that the method still has limitations. Some resolutions were supported by bootstrap analysis both in

parsimony and distance analysis (C, D and O; F and H; A and B). Bremer support values also suggest that, while there is sufficient information to resolve the genetic relationships, the topology requires stronger support. For general aspects of genetic distances, this level of precision is acceptable as an expedite criteria. Therefore, it is viable to gather basic genetic information for *Calophyllum brasiliense* using dominant markers, particularly with RAPD technique and even at small geographic scales. More precise surveys would require larger datasets or markers with higher discriminatory power.

The parsimony analysis indicated a north-south relationship. Group A joins G (north) with F (intermediate) and H (south). The same pattern is apparent for the other groups. The natural dispersion of *C. brasiliense* is mainly through bats (Marques and Joly 2000). However, the observed pattern may also be a consequence of a major road (BR101) that runs is near the sampling site (ca. 200m) in a north/south pattern. Such a large road may have influenced bat ecology, limiting or restricting the pattern of seed dispersal.

A strategy of sampling to compose seed orchard can be proposed based in the unrooted parsimony tree. This is far from being the only option, but has the advantage of minimizing costs and maximizing genetic diversity. We propose sampling most divergent genotypes detected in tree topology. Obviously, the small scale sampling allows inclusion of all 17 samples to a seed orchard. This would be satisfactory to capture the inter-clade diversity and most of the population diversity. Nonetheless, a larger number of individuals would be needed to adequately represent diversity within clades and maximize the probability of inclusion of rare alleles.

## CONCLUSIONS

Using dominant markers, particularly with the RAPD technique, seems a viable strategy to collect basic genetic information for *Calophyllum brasiliense*, even at small geographic scales. This method revealed high genetic diversity in the studied population of *C. brasiliense*, a typical feature of native uncultivated tree species.

Although lacking strong support, it is possible to explore genetic distances and recognize patterns of distribution of the studied genotypes with

RAPD. This information allows the development of a sampling strategy to ensure maximum genetic representation when selecting representatives for a seed orchard.

## REFERENCES

Bremer K. Branch support and tree stability. *Cladistics*. 1994; 10: 295-304.

Caldas L. Pomares de sementes de espécies nativas – As funções das redes de sementes. In: Higa AR. and Silva LD. Pomar de sementes de espécies florestais nativas. 1st ed. FUPEF; 2006. 227-242.

Cheung WY, Hubert N, Landry BS. A simple and rapid DNA microextraction method for plant, animal, and insect suitable for RAPD and other PCR analyses. *Genome Res*. 1993; 3: 69-70.

Cruz CD. Programa Genes: Biometria. Viçosa (MG): Editora UFV; 2006.

Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a no metric hypothesis. *Psychometrika*. , 1964; 29(1): 1-27.

Lazar I. [homepage on the internet].Gel Analyzer. Debrecen, Hungary: University of Debrecen; updated in 2010; accessed in 2013 Jan 11. Available from: <http://www.gelanalyzer.com/index.html>

Manly BFJ. Randomization, bootstrap and Monte Carlo methods in Biology. London: Chapman & Hall, 1997.

Marques MCM, Joly CA. Estrutura e dinâmica de uma população de *Calophyllum brasiliense* Camb. em floresta higrófila do sudeste do Brasil. *Rev Bras Bot*. 2000; 23(1):107-12.

Muchugi A, Kadu C, Kindt R, Kipruto H, Lemurt, S, Olale K, Nyadoi P, Dawson I, Jamnadass, R. Molecular markers for tropical trees: A Practical guide to principles and procedures. ICRAF Technical Manual no. 9. World Agroforestry Centre Nairobi; 2009.

Nei M, Li W-H. Mathematical model for studying genetic variation in terms of restriction

endonucleases. *Proc. Natl. Acad. Sci. USA*. 1979; 76(10): 5269-73.

Nei M. Genetic distance between populations. *Am. Nat.* 1972; 106:283-92.

Newbury HJ, Ford-Lloyd BV. The use of RAPD for assessing variation in plants. *Plant Growth Regul.* 1993; 12:43-51, 1993.

Oliveira-Filho AT, Ratter JA. A study of the origin of central brazilian forests by the analysis of plant species distribution patterns. *Edinb J Bot.* 1995; 52: 141-94.

Reitz R, Klein RM, Reis A. Projeto madeira de Santa Catarina. *Sellowia*. 1978; 28(30): 218-24.

Rohlf FJ. NTSYS-pc v. 1.8. Numerical taxonomy and multivariate analysis system. Setauket, NY: Applied Biostatistics Inc. 1993.

Sebbenn, AM. Número de árvores matrizes e conceitos genéticos na coleta de sementes para reflorestamentos com espécies nativas. *Rev Inst Flor.* 2002; 14: 115-132.

Sebbenn, AM. Sistemas de reprodução em espécies arbóreas tropicais e suas implicações para a seleção de árvores matrizes para reflorestamentos ambientais. In: Higa AR. and Silva LD. Pomar de sementes de espécies florestais nativas. 1st ed. FUPEF; 2006. 93-138.

Swofford, D. L. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4b10. Sunderland, Massachusetts: Sinauer Associates. 2003.

Tivang JG, Nienhuis J, Smith OS. Estimation of sampling variance of molecular marker data using the bootstrap procedure. **Theor Appl Genet.** 1994; 89(25):264-9.

Yeh F, Yang RC, Mao J, Ye Z, Boyle TJB. The user-friendly shareware for Population Genetic Analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, 1999.

Recebido: 04/05/2013  
Received: 05/04/2013

Aprovado: 19/09/2013  
Approved: 09/19/2013