RESEARCH

# Molecular Markers Reveal Limited Genetic Diversity in a Large Germplasm Collection of the Biofuel Crop *Jatropha curcas* L. in Brazil

Tatiana B. Rosado, Bruno G. Laviola, Danielle A. Faria,
Marilia R. Pappas, Leonardo L. Bhering, Betania Quirino, and Dario Grattapaglia*

**ABSTRACT**

The genetic diversity of a comprehensive germplasm collection involving 192 *Jatropha curcas* L. accessions collected throughout Brazil, spanning a wide latitudinal range from the states of Maranhão (1°49' S, 44°52' W) to Rio Grande do Sul (29°33' S, 55°07' W), was studied with 96 random amplified polymorphic DNA (RAPD) primers and six selected microsatellite markers. Only 23 of the 381 replicated RAPD markers and one microsatellite were polymorphic. Surprisingly, all accessions were homozygous at all but one microsatellite, in contrast with the outcrossing mating system reported for the species, suggesting that *J. curcas* not only supports selfing but possibly breeds by geitonogamy. Similarity based clustering revealed only 43 unique multilocus profiles in the 192 accessions. The probabilities of accessions with indistinguishable multilocus profiles being true duplicates varied between 83 and 99%. No relationship between clustering of accessions and geographic origin was observed, suggesting that *J. curcas* has experienced a widespread dispersion across regions by seeds and possibly vegetative propagules. The narrow genetic base and extent of potentially duplicated accessions likely reflects a recent common ancestry, drift, and intensive selection of the currently cultivated materials since the time of introduction. This result highlights an urgent need for the introduction of new and diverse accessions to this germplasm collection if Brazil is to drive and sustain successful breeding programs.

T.B. Rosado, B.G. Laviola, L.L. Bhering, and B.F. Quirino, Embrapa Agroenergy, PqEB, Brasilia, 70770-901 DF, Brazil; D.A. Faria, M.R. Pappas, and D. Grattapaglia, EMBRAPA Genetic Resources and Biotechnology, PqEB, Brasília, 70770-970 DF, Brazil; D. Grattapaglia and B.F. Quirino, Graduate Program in Genomic Sciences and Biotechnology, Universidade Catolica de Brasília–SGAN 916 modulo B, Brasília 70790-160 DF, Brazil. Received 25 Feb. 2010. *Corresponding author: (dario@cenargen.embrapa.br).

**Abbreviations:** AFLP, amplified fragment length polymorphism; DI, diversity index; EPAMIG, Empresa de Pesquisa Agropecuaria de Minas Gerais; GS, genetic similarity; $I_b$, band informativeness; ISSR, inter-simple sequence repeat; PCR, polymerase chain reaction; $P_{DUP}$, probability of duplication; $P_{MP}$, probability of occurrence of the multilocus profile; RAPD, random amplified polymorphic DNA; Rp, resolving power; SSR, simple sequence repeat; $T_a$, annealing temperature; UFLA, University of Lavras; UPGMA, unweighted pair group method using arithmetic averages.

*JATROPHA CURCAS* L. is a perennial shrub of the Euphorbiaceae family widely dispersed throughout the world and thought to be native to Central America (Fairless, 2007). A renewed interest in finding alternatives to petroleum-based fuels has placed *J. curcas* seed oil as one of the promising sources of biodiesel (Openshaw, 2000). *Jatropha curcas* oil is relatively simple to convert to biodiesel by chemical (Berchmans and Hirata, 2008) or biological transesterification reactions (Modi et al., 2007). Its adaptation to different environments, ability to grow in marginal soils, drought tolerance, and insect and disease resistance make *J. curcas* attractive for cultivation (Fairless, 2007). In Brazil, *J. curcas* was introduced centuries ago. Today, *J. curcas* is planted throughout Brazil as isolated plants or in small groups as living fences in backyards and

Published in Crop Sci. 50:2372–2382 (2010).
doi: 10.2135/cropsci2010.02.0112
Published online 27 Sept. 2010.

small farms. This indicates a strong human action in maintaining and disseminating this crop. The first large scale, systematic *J. curcas* plantations were established in Brazil in the 80s for academic studies and commercial purposes (Saturnino et al., 2005). Current estimates indicate that 40,000 ha of *J. curcas* are planted in Brazil (Mike Lu, personal communication, 2010) and the current trend is that this planted area will increase in coming years. Trees produce seeds within 1 yr, have reproductive life cycle of over 30 yr, and can produce up to 1300 L of oil $ha^{-1}$ (Fairless, 2007; Francis et al., 2005; Openshaw, 2000).

In spite of increasingly widespread interest in planting *J. curcas,* very little is known about the genetic diversity of available germplasm resources. Furthermore, systematic breeding efforts are still in their infancy and currently there are no agronomically elite cultivars of *J. curcas* available (Carels, 2009). *Jatropha curcas* is a monoecious, predominantly outcrossing species but is fully self-compatible and subject to geitonogamy (Ginwal et al., 2005; Liu et al., 2008; Luo et al., 2007) with a genome size of 416 Mb distributed in 11 pairs of chromosomes ($2n = 22$) (Abdelgadir et al., 2009; Carvalho et al., 2008). Its preferential mating system would, in principle, suggest a relatively high degree of genetic variation, which is a key element to start a promising breeding program. Nevertheless, several studies have reported low genetic diversity in the relatively limited germplasm collections of *J. curcas* surveyed to date, typically including between 40 and 70 accessions using isozyme, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR), and microsatellites (Basha and Sujatha, 2007; Basha et al., 2009; Kumar et al., 2009a, b; Ram et al., 2008; Sun et al., 2008; Tatikonda et al., 2009). To date, these marker classes have been useful to evaluate, at the genome-wide level, the extent of genetic diversity in *J. curcas*, although more powerful molecular tools will likely be developed once the recently sequenced genome for this crop (http://www.syntheticgenomics.com/media/press/52009.html; verified 18 July 2010) becomes publicly available.

A collection of 192 accessions was assembled between 2006 and 2009 by sampling expeditions performed by EMBRAPA (Empresa Brasileira de Pesquisa Agropecuaria) in several locations throughout Brazil, covering areas from the equatorial Maranhão state (1°49' S lat) to the southern state of Rio Grande do Sul (30° S lat). This existing set of accessions effectively represents the currently adapted planting stock in the country and constitutes an important initial resource to begin breeding efforts involving the recurrent selection of elite genotypes to be used as parents in seed production areas or vegetatively propagated for operational deployment. In this study we assessed the molecular diversity of this collection to: (i) describe the extent of molecular diversity available in the germplasm that the country currently holds; (ii) verify possible patterns of genetic diversity that could be explained by geographical origin so as to direct further collection efforts; and (iii) detect probable duplicate accessions. This knowledge would provide fundamental information about this germplasm to drive additional collections and potentially start systematic breeding efforts. To examine the genetic diversity and structure of this germplasm collection we used both a selected set of robust and fully replicated RAPD markers and a panel of microsatellites that have shown a high degree of multiallelism in previous studies. By using two complementary marker technologies, our goal was to better capture the existing diversity by simultaneously supplying wider genome coverage with the multiplexing ability of RAPD markers and codominant marker information provided by the microsatellites.

# MATERIALS AND METHODS

## Plant Material and DNA Extraction

*Jatropha curcas* is currently found throughout the Brazilian territory from coastal regions to the inland at altitudes up to 1200 m. It is typically found in a dispersed fashion associated with human intervention. A set of 192 accessions of *J. curcas* collected in different geographical regions throughout Brazil were therefore surveyed (Fig. 1; Supplementary Material S1). Each accession was coded by a sequential number, followed by three letters denoting the collection site or germplasm bank, followed by the two letter code for the Brazilian state. Out of the 192 accessions, 29 were provided directly by two germplasm banks: eight accessions (#52 through #59) from the EPAMIG (Empresa de Pesquisa Agropecuaria de Minas Gerais) bank in Nova Porteirinha (three letter code NPO) and 21 accessions (#88 through #108) from the UFLA (University of Lavras) bank (three letter code Lav). For the remaining 163 accessions, seeds were collected from individual plants typically found in backyards, gardens, and abandoned pastures or as hedgerows as well as on small scale plantations. An accession was defined as a single plant wherefrom seeds were collected and that was distant at least 500 m from the next collected accession. This procedure was adopted to minimize relatedness among accessions based on the assumption that neighboring plants might have a recent common ancestor. Seeds of these accessions were planted as a common garden in the germplasm bank of Embrapa Agroenergy in Brasilia, Federal District, Brazil. Total genomic DNA was extracted from the younger leaves of two individual plants for each accession, constituting two biological replicates, using a CTAB protocol (Doyle and Doyle, 1990) with minor modifications as described earlier (Grattapaglia and Sederoff, 1994). Briefly, 150 g of fresh leaf tissue were ground directly in 700 μL of extraction buffer (2% CTAB, 20 mM EDTA, 2% PVP, 1.4 M NaCl, 100 mM Tris-HCl pH 8.0, and 1% β-mercaptoethanol) using a Tissuelyser (Qiagen, Alameda, CA) and samples incubated at 65°C for 1 h. The supernatant was extracted twice with 500 μL of chloroform:isoamylalcohol (24:1 v/v) and the DNA was precipitated with 1 volume of isopropanol and washed with 70% ethanol overnight and once with 95% ethanol. The pelleted DNA was air dried and resuspended in 100 μL of TE buffer containing RNase A (10 mg $ml^{-1}$) and stored at −20°C. The DNA concentration was measured using a Nanodrop spectrophotometer (Nanodrop Product, Wilmington, DE) and the concentration of each sample was adjusted to 2 to 5 ng $\mu L^{-1}$.


Figure 1. Geographic distribution of the sites where the 192 *Jatropha curcas* accessions were collected. Detailed information regarding the municipality, state, and exact global positioning system (GPS) coordinates for the collection sites indicated by dots are listed in Supplementary Material S1.

## RAPD Marker Analysis

Amplification, interpretation, and robustness analysis of RAPD markers was performed as described earlier (Grattapaglia and Sederoff, 1994; Reis and Grattapaglia, 2004). For each accession two biological replicate samples were analyzed and only those RAPD markers that consistently amplified in both replicates were scored. A total of 96 decamer arbitrary sequence primers (Operon Technologies, Inc., Huntsville, AL; DT Technologies, Coralville, IA) were used, 48 of which were selected based on previous studies that reported primers amplifying a larger number of polymorphic markers for *J. curcas* (Basha and Sujatha, 2007; Gupta et al., 2008; Kumar et al., 2009b; Pamidimarri et al., 2009b; Ram et al., 2008; Subramanyam et al., 2009). Each polymerase chain reaction (PCR) was performed in a total volume of 13 μL containing 1.3 μL of 10x PCR buffer (10 mM Tris pH 9.0, 50 mM KCl, and 1.5 mM $MgCl_2$), 1.3 μL of 2.5 mM each dNTP, 0.4 μL $MgCl_2$ (50 mM), 2 μL primer (10 ng $\mu L^{-1}$), 0.2 μL Taq DNA polymerase (5 U $\mu L^{-1}$) (Phoneutria; Belo Horizonte, Minas Gerais, Brazil), 2.0 μL template DNA (5 ng $\mu L^{-1}$), and 5.8 μL RNase-free water. Polymerase chain reaction amplification was performed in a thermal cycler (Eppendorf Mastercycler ep gradient S; Eppendorf AG, Hamburg, Germany) with the following program: an initial denaturation step at 94°C for 3 min, 42 cycles of denaturation at 94°C for 30 s, primer annealing at 32°C for 1 min, extension at 72°C for 2.5 min, and a final extension at 72°C for 4 min. Polymerase chain reaction products were separated on 1.5% agarose gel in 1x SBE buffer by electrophoresis at 120 V for 3 h and visualized after staining with ethidium bromide.

## Microsatellite Marker Analysis

Six microsatelllite markers for *J. curcas* developed by Pamidimarri et al. (2009a) were selected based on high genetic information

content and allele size range for multiplexing. These microsatellites were amplified by PCR with the forward primer of each microsatellite labeled with fluorochrome 6-FAM. Loci were combined according to the expected allele size range so as to allow simultaneous amplification and electrophoretic separation. The multiplex systems used were a 3-plex composed by markers jcps6, jps9, jds10 (annealing temperature $T_a$ = 56°C) and a 2-plex of markers jcds24 and jcds41 ($T_a$ = 44°C). Microsatellite jcps20 was amplified alone ($T_a$ = 56°C) as it did not perform adequately in any of the multiplex systems. The PCR amplification reaction (5 μL) was performed using the Multiplex PCR Kit (Qiagen, Alameda, CA) containing 2.5 μL of 10x PCR Master Mix, 0.5 μL de Q-Solution, 0.1 μL of each primer at a concentration of 10 μM (final concentration of 0.2 μM), 0.4 μL RNase-free water, and 1.5 μL template DNA (2 ng $\mu L^{-1}$). Polymerase chain reaction amplification was performed in a thermal cycler (Eppendorf Mastercycler ep gradient S; Eppendorf AG, Hamburg, Germany) with the following program: an initial denaturation step at 95°C for 15 min, 35 cycles of denaturation at 95°C for 30 s, primer annealing for 90s at the temperature described above specific to each multiplex system, extension at 72°C for 1 min, followed by a final extension at 72°C for 30 s. Polymerase chain reaction amplified products were electroinjected in an ABI PRISM 3100XL Genetic Analyzer (Applied Biosystems, Foster City, CA) and data was collected under virtual filter D using GeneScan 2.0 and alleles declared using Genotyper 2.1 (Applied Biosystems, Foster City, CA). An internal size standard labeled with ROX (Brondani and Grattapaglia, 2001) was used for sizing alleles. The software AlleloBin was used to classify observed microsatellite allele sizes into representative discrete alleles using a least-square minimization algorithm (Idury and Cardon, 1997).

## Data Analysis

Polymorphic RAPD markers were visually scored in a binary fashion for presence or absence of bands. For each polymorphic RAPD marker, binary RAPD marker frequencies [presence (P) or absence (1 – P)] were estimated from the genotypic data of the 192 accessions. A diversity index (DI) and band informativeness ($I_b$) were estimated for each polymorphic RAPD band and the resolving power (Rp) estimated for the RAPD primer as described by (Prevost and Wilkinson, 1999). Monomorphic RAPD markers were scored but were not used in the genetic distance estimation as they did not provide informative data. Microsatellite markers were scored and genotypes used together with the binary RAPD data in the distance analyses. A Jaccard's similarity coefficient (Jaccard, 1901) was used to estimate genetic similarity (GS) between pairs of genotypes according to the equation: $GS_{ij} = a/(a + b + c)$, where $GS_{ij}$ is the genetic similarity measurement between individuals *i* and *j*, *a* is the number of polymorphic markers (RAPD bands or microsatellite alleles) present in both individuals, and *b* and *c* are the number of bands present only in individuals *i* and *j*, respectively, and not in their counterparts. The GS matrix was used to perform cluster analysis with the unweighted pair group method using arithmetic averages (UPGMA) algorithm (Sneath and Sokal, 1973). These analyses were performed using modules in the NTSYS-PC software, version 2.1 (Rohlf, 2002). The approximate probability that two or more accessions with indistinguishable multilocus marker profiles were true germplasm duplicates ($P_{DUP}$ = probability of duplication) and not indistinguishable due to chance was calculated following the estimation of the random match probability of multilocus genotypes (Evett and Weir, 1998), also called probability of identity, commonly used for germplasm analysis (Besnard et al., 2001; Gorg et al., 1992; Irisha et al., 2010). Briefly, frequencies of RAPD marker or microsatellite allele observed in a multilocus profile were used to estimate the probability of occurrence of the multilocus profile ($P_{MP}$) in this germplasm collection by using the product rule in the expression:

$$P_{MP} = \Pi\, p_{ij}$$

where $p_{ij}$ is the frequency of the *j*th allele at the *i*th RAPD or microsatellite marker. As this profile frequency corresponds to the random match probability between two unrelated accessions, the probability of a true match, that is, that the two accessions are in fact the same and thus a duplicate, is equal to its complement:

$$P_{DUP} = 1 - P_{MP}$$

This computation is approximate as it assumes independence between markers in the profile (Evett and Weir, 1998), a condition not tested in this study.

# RESULTS

## RAPD and Microsatellite Markers

Ninety-six RAPD primers amplified 381 robust and easily interpretable markers supplying on average almost four clearly interpretable markers per RAPD primer. However, only 23 of the 381 were polymorphic (6.2%) across the 192 accessions (Supplementary Material S2). These 23 informative genomic loci were amplified with 12 primers (Table 1), thus providing an average of 1.9 informative loci per selected primer with a range of one to a maximum of four. The RAPD amplicon size ranged from 300 to 2500 bp. With a few exceptions, for most RAPD markers the minimum allele frequency either for the band present (P) or absent (1 – P) was low, between 1 and 5%, resulting in a low information content for discrimination purposes as indicated by the DI and $I_b$ estimates (Table 1). Random amplified polymorphic DNA markers OPV_18_600, OPT_14_1050, OPN_19_1500, and OPA_11_1050 were the most informative with the frequencies of the alternative alleles closer in value and therefore the highest DI and $I_b$. Consequently, the Rp were also the highest for the corresponding RAPD primers (Table 1; Supplementary Material S2).

All six microsatellites genotyped amplified alleles within the expected size range. However, only one or two alleles were observed at each one of the six loci, with a total of only eight different alleles in all the 192 germplasm accessions analyzed in spite of the expected multiallelic nature of these hypervariable markers (Table 2; Supplementary Material S3). Four microsatellites were monomorphic, displaying a single allele across all the 192 *J. curcas* accessions (Table 2). Microsatellite jcps6 was also monomorphic but revealing what seemed to be a heterozygous genotype in all the accessions (Genotype 269/285 bp),

**Table 1. Summary of the 23 polymorphic random amplified polymorphic DNA (RAPD) markers amplified with 12 arbitrary sequence primers in the *Jatropha curcas* germplasm collection.**

| RAPD marker | Primer sequence (5' to 3') | *n*† | P‡ | (1 – P)§ | DI ¶ | $I_b$# | $R_p$†† |
|---|---|---|---|---|---|---|---|
| OPR_19_750 | CCTCCTCATC | 192 | 0.016 | 0.984 | 0.031 | 0.031 | 0.031 |
| OPV_18_1100 | TGGTGGCGTT | 192 | 0.948 | 0.052 | 0.099 | 0.104 | – |
| OPV_18_600 | TGGTGGCGTT | 188 | 0.441 | 0.559 | 0.493 | 0.883 | 0.987 |
| OPP_01_900 | GTAGCACTCC | 192 | 0.932 | 0.068 | 0.126 | 0.135 | – |
| OPP_01_850 | GTAGCACTCC | 192 | 0.031 | 0.969 | 0.061 | 0.063 | – |
| OPP_01_450 | GTAGCACTCC | 192 | 0.031 | 0.969 | 0.061 | 0.063 | 0.260 |
| OPT_14_1050 | AATGCCGCAG | 174 | 0.736 | 0.264 | 0.389 | 0.529 | – |
| OPT_14_1200 | AATGCCGCAG | 192 | 0.026 | 0.974 | 0.051 | 0.052 | – |
| OPT_14_950 | AATGCCGCAG | 192 | 0.010 | 0.990 | 0.021 | 0.021 | – |
| OPT_14_800 | AATGCCGCAG | 192 | 0.036 | 0.964 | 0.070 | 0.073 | 0.675 |
| OPJ_20_700 | AAGCGGCCTC | 192 | 0.979 | 0.021 | 0.041 | 0.042 | – |
| OPJ_20_850 | AAGCGGCCTC | 192 | 0.021 | 0.979 | 0.041 | 0.042 | 0.083 |
| OPN_19_1500 | GTCCGTACTG | 176 | 0.756 | 0.244 | 0.369 | 0.489 | 0.489 |
| OPV_14_1050 | AGATCCCGCC | 192 | 0.021 | 0.979 | 0.041 | 0.042 | – |
| OPV_14_600 | AGATCCCGCC | 192 | 0.021 | 0.979 | 0.041 | 0.042 | 0.083 |
| OPA_11_1600 | CAATCGCCGT | 192 | 0.031 | 0.969 | 0.061 | 0.063 | – |
| OPA_11_1050 | CAATCGCCGT | 192 | 0.266 | 0.734 | 0.390 | 0.531 | – |
| OPA_11_400 | CAATCGCCGT | 192 | 0.021 | 0.979 | 0.041 | 0.042 | 0.635 |
| OPC_18_700 | TGAGTGGGTG | 192 | 0.021 | 0.979 | 0.041 | 0.042 | – |
| OPC_18_650 | TGAGTGGGTG | 192 | 0.036 | 0.964 | 0.070 | 0.073 | 0.115 |
| OPK_12_900 | TGGCCCTCAC | 192 | 0.016 | 0.984 | 0.031 | 0.031 | 0.031 |
| OPC_14_700 | TGCGTGCTTG | 192 | 0.932 | 0.068 | 0.126 | 0.135 | 0.135 |
| OPC_10_1800 | TGTCTGGGTG | 192 | 0.016 | 0.984 | 0.031 | 0.031 | 0.031 |

†*n*, number of accessions genotyped with full replication.

‡P, frequency of the presence of RAPD band.

§(1 – P), frequency of the absence of RAPD band.

¶DI, Diversity index, estimated as DI = $\Sigma[p2(1 - P)2]$.

#Ib, RAPD band informativeness (see text for details).

††Rp, RAPD primers resolving power (see text for details).

although possibly reflecting a duplicated locus and not a true heterozygous genotype (Supplementary Material S3). Marker jcps20 was the only microsatellite that could be classified as polymorphic, although borderline. It only displayed two alleles, one of them (209 bp) with a very low frequency of 0.0054, and only nine accessions were heterozygous (209/216) (Table 2; Supplementary Material S3). Therefore only data for locus jcps20 was included in the distance analysis although with little overall contribution.

## Analysis of Genetic Distance and Duplicated Accessions

Jaccard's genetic similarity coefficient varied from 0.14 to 1.0 among the 192 accessions and a high average similarity of 0.89 was estimated. The lowest genetic similarity (highest genetic divergence) was observed when accession 65-Jaí-MG was compared to accessions 1-RiV-GO and 153-Ana-Go, with similarities of 0.11 and 0.35, respectively. Complete multilocus similarity

**Table 2. Microsatellite markers used in the study and results of alleles observed (in bp) with their respective frequencies.**

| Marker (GenBank) | Primer sequences (5' to 3') | Repeat motif | $T_a$† (°C) | Allele (frequency) | Allele (frequency) | $H_{obs}$‡ |
|---|---|---|---|---|---|---|
| jcps06 (EU586346) | F:CCAGAAGTAGAATTATAAATTAAA<br>R:AGCGGCTCTGACATTATGTAC | (AT)3G(TA)3…(CT)3…<br>(GT)5CT(GT)3 | 44.0 | 269 (0.50) | 285 (0.50) | 1.00 |
| jcps09 (EU586347) | F:GTACTTAGATCTCTTGTAACTAACAG<br>R:TATCTCTTGTTCAGAAATGGAT | (GT)3GC(TG)2A(GT)3 | 48.0 | 142 (1.00) | – | 0.00 |
| jcps10 (EU586340) | F:CATCAAATGCTAATGAAAGTACA<br>R:CACACCTAGCAAACTACTTGCA | (TG)6CACGCA(TG)4 | 46.5 | 109 (1.00) | – | 0.00 |
| jcps20 (EU586348) | F:ACAGCAAGTGCACAACAATCTCA<br>R:TACTGCAGATGGATGGCATGA | (TG)12(GA)22 | 55.0 | 209 (0.027) | 216 (0.973) | 0.0054 |
| jcps24 (EU586341) | F:GGATATGAAGTTTCATGGGACAAG<br>R:TTCATTGAATGGATGGTTGTAAGG | (CA)5(TA)8(CA)4…<br>(TA)3GA(TA)4 | 51.0 | 200 (1.00) | – | 0.00 |
| jcps41 (EU586342) | F: AACACACCATGGGCCACAGGT<br>R:TGCATGTGTGCGGGTTTGATTAC | (CA)6(TA)2 | 56.6 | 108 (1.00) | – | 0.00 |

†$T_a$, annealing temperature.

‡$H_{obs}$, observed heterozygosity.

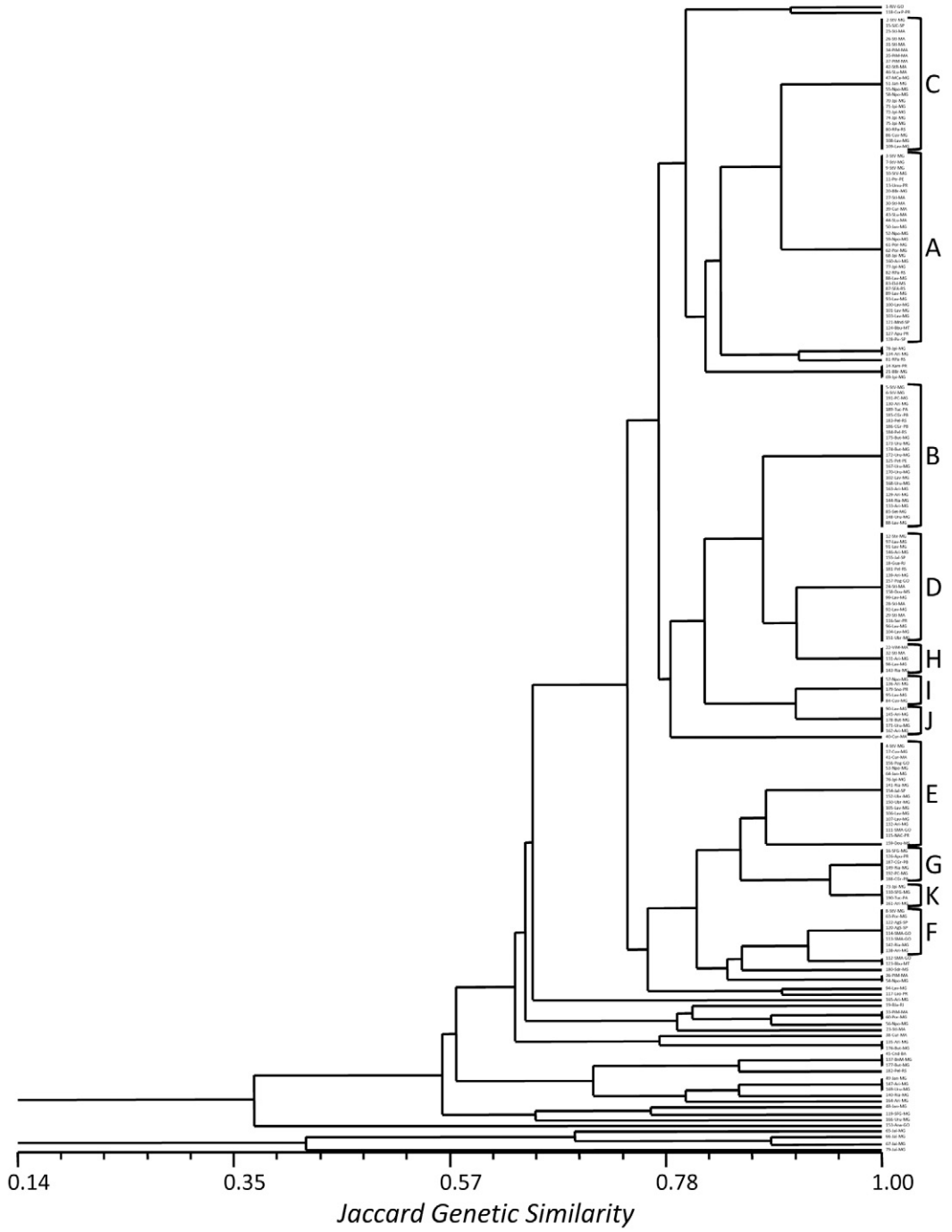


Figure 2. Unweighted pair group method using arithmetic averages (UPGMA) phenogram representing the genetic relationships among 192 accessions of *J. curcas* collected in Brazil based on a Jaccard genetic similarity matrix obtained using a multilocus allelic profile of 23 polymorphic random amplified polymorphic DNA (RAPD) and one microsatellite. Accessions with indistinguishable multilocus marker genotypes (Jaccard similarity equal to 1.0) were grouped in letter coded profiles to assess the probability of being true duplicates. Profiles that had four or more indistinguishable accessions are indicated by letters and delimiting brackets (see text for more detail).

was observed among several accessions. However, no clear pattern of clustering among accessions was observed that correlated to their geographical origin. In some cases accessions with indistinguishable profiles were collected in very distant states (denoted by the two last letters in the accession code) and in others accessions from the same collection site clustered in different groups (Fig. 2).

Accessions with indistinguishable multilocus marker genotypes were grouped in a set of profiles (Fig. 2; Table 3). Profile (A) is the largest group with 33 entries. It includes accessions collected in eight different states spanning a very wide latitudinal range from Cururupu-Maranhão (1°49' S, 44°52' W) to São Francisco de Assis-Rio Grande do Sul (29°33' S, 55°07' W) locations over 3000 km apart. The same pattern was observed for other four groups that had

**Table 3. Groups of *J. curcas* accessions with indistinguishable multilocus profiles (letter coded) with their respective probabilities of occurrence ($P_{MP}$) in the germplasm collection and the probability with which the accessions in each group are true duplicates ($P_{DUP}$).**

| Profile | No. Accessions | $P_{MP}$ | $P_{DUP}$ | Accessions |
|---|---|---|---|---|
| A | 33 | 0.1306 | 0.8694 | 3-StV-MG; 7-StV-MG; 9-StV-MG; 10-StV-MG; 11-Ptr-PE; 13-Umu-PR; 20-BBr-MG; 27-StI-MA; 30-StI-MA; 39-Cur-MA; 43-SLu-MA; 44-SLu-MA; 50-Jan-MG; 52-Npo-MG; 59-Npo-MG; 61-Por-MG; 62-Por-MG; 68-Jpi-MG; 160-Ari-MG; 77-Jpi-MG; 82-RPa-RS 88-Lav-MG; 83-Eld-MS; 87-SFA-RS; 89-Lav-MG; 93-Lav-MG; 100-Lav-MG; 101-Lav-MG; 103-Lav-MG; 121-Mnd-SP; 124-Bbu-MT; 127-Apu-PR; 128-Pir-SP |
| B | 25 | 0.1652 | 0.8348 | 5-StV-MG; 6-StV-MG; 191-PC-MG; 130-Ari-MG; 189-Tuc-PA; 185-CGr-PB; 183-Pel-RS; 186-CGr-PB; 184-Pel-RS; 175-But-MG; 173-Uru-MG; 174-But-MG; 172-Uru-MG; 125-Pet-PE; 167-Uru-MG; 170-Uru-MG; 102-Lav-MG; 168-Uru-MG; 163-Ari-MG; 129-Ari-MG; 144- Ria-MG; 133-Ari-MG; 85-Set-MG; 148-Uru-MG; 88-Lav-MG; |
| C | 23 | 0.0311 | 0.9689 | 2-StV-MG; 15-SJC-SP; 25-STI-MA; 26-StI-MA; 31-StI-MA; 34-PiM-MA; 35-PiM-MA; 37-PiM-MA; 42-StR-MA; 47-MCa-MG; 51-Jan-MG; 55-Npo-MG; 58-Npo-MG; 70-Jpi-MG; 71-Jpi-MG; 72-Jpi-MG; 74-Jpi-MG; 75-Jpi-MG; 80-RPa-RS; 86-Cuv-MG; 108-Lav-MG; 109-Lav-MG; |
| D | 19 | 0.1215 | 0.8785 | 12-Ste-MG; 97-Lav-MG; 91-Lav-MG; 146-Ari-MG; 155-JAL-SP; 18-Gua-RJ; 181-Pel-RS; 139-ARI-MG; 157-POG-GO; 23-STI-MA; 158-DOU-MS; 99-Lav-MG; 23-STI-MA; 92-Lav-MG; 23-STI-MA; 116-SER-PR; 96-Lav-MG; 104-Lav-MG; 151-UBR-MG; |
| E | 17 | 0.0439 | 0.9561 | 4-STV-MG; 17-Cuv-MG; 41-Cur-MA; 156-POG-GO; 53-NPO-MG; 64-JAN-MG; 76-JPI-MG; 141-Ria-MG; 154-JAL-SP; 152-UBR-MG; 150-UBR-MG; 105-Lav-MG; 106-Lav-MG; 107-Lav-MG; 132-Ari-MG; 111-SMA-GO; 115-NAC-PR; |
| F | 8 | 0.0347 | 0.9653 | 8-STV-MG; 63-Por-MG; 122-AGS-SP; 120-AGS-SP; 114-SMA-GO; 113-SMA-GO; 142-Ria-MG; 138-Ari-MG |
| G | 6 | 0.0158 | 0.9842 | 16-SFG-MG; 126-APU-PR; 187-CGR-PB; 149-Ria-MG; 192-PC-MG; 188-CGR-PB; |
| H | 5 | 0.0070 | 0.9930 | 22-VIM-MA; 23-STI-MA; 131-ARI-MG; 98-Lav-MG; 143-Ria-MG |
| I | 5 | 0.0393 | 0.9607 | 57-NPO-MG; 136-Ari-MG; 179-SNO-PR; 95-Lav-MG; 84-Cuv-MG |
| J | 5 | 0.0534 | 0.9466 | 90-Lav-MG; 145-Ari-MG; 178-But-MG; 171-Uru-MG; 162-Ari-MG |
| K | 4 | 0.0051 | 0.9949 | 73-JPI-MG; 110-SFG-MG; 190-TUC-PA; 161-Ari-MG; |
| L | 3 | 0.0010 | 0.9990 | 14-Xam-PR; 21-BBr-MG; 69-Jpi-MG |
| M | 3 | 0.0002 | 0.9998 | 45-CND-BA; 137-BnM-MG; 177-But-MG |
| N | 3 | 0.0002 | 0.9998 | 49-Jan-MG; 147-Ari-MG; 169-Uru-MG; |
| O | 2 | 0.0457 | 0.9543 | 78-Jpi-MG; 134-Ari-MG; |
| P | 2 | 0.0499 | 0.9501 | 122-SMA-GO; 123-BBU-MT; |
| Q | 2 | 0.0119 | 0.9881 | 36-PIM-MA; 54-NPO-MG |
| R | 2 | 0.0055 | 0.9945 | 33-PIM-MA; 60-Por-MG |
| S | 2 | 0.0009 | 0.9991 | 135-Ari-MG; 176-BU-MG |

more than ten indistinguishable accessions at the molecular level but collected at a geographic distance of over 3000 km (Table 3). On the other hand, accessions collected in the same or very close localities had distinguishable marker profiles and clustered in different groups such as those from João Pinheiro–MG (Jpi), Janaúba–MG (Jan), and Arinos–MG (Ari). This same pattern was striking for the eight accessions collected in three municipalities of the state of Rio Grande do Sul (Rio Pardo, São Francisco de Assis e Pelotas). In spite of their geographic proximity, these eight accessions had six distinct profiles: profile (A) (82-RPa-RS and 87-SFA-RS); profile (B) (183-Pel-RS and 184-Pel-RS); profile (C) (80-RPa-RS); profile (D) (181-Pel-RS); and accessions 81-RPa-RS and 182-Pel-RS with unique profiles. The genetic distance analysis also showed a markedly separate cluster with four accessions (65, 66, 67, and 79) (Fig. 2) that surprisingly were all derived from the same commercial plantation in Jaiba, state of Minas Gerais (Supplementary Material S1). Additionally a single accession (153), also derived from a commercial plantation in Anapolis, state of Goias, also showed a substantial genetic divergence from the main cluster of accessions.

Clustering of genotypes by the UPGMA method and representing them in a phenogram revealed 43 distinct multilocus profiles for the 192 accessions, evidenced by the number of final branches at similarity equal to one in the phenogram (Fig. 2). This result indicated that, at the molecular level, this collection involves only 43 distinguishable multilocus genotypes with this set of RAPD and microsatellite markers. Accessions 66-Jaí-MG, 67-Jaí-MG, 65-Jaí-MG, 79-Jaí-MG, and 153-Ana-GO were considerably differentiated from the remainder (Fig. 2). Out of these 43 multilocus profiles, 24 were observed in a single accession and 19 were seen for two or more accessions. These 19 multilocus profiles were coded with letters sequentially with the increasing number of accessions and their probability of occurrence ($P_{MP}$) estimated (Table 3). The complement of this probability corresponds to the probability of the accessions that belong to the same group being true duplicated accessions ($P_{DUP}$) (Table 3). So, for example, the

probability of the 33 accessions with profile (A) being truly duplicated germplasm is 1 – 0.1306 = 0.8694 or 86.94%; for the 25 accessions with profile (B) this probability amounts to 83.48%; and for profile (C), with 23 accessions, the probability is 96.89%. As the number of accessions with the same profile gets smaller, the probability of being true duplications gets larger due to the presence of rarer alleles in the profile that contribute to a lower $P_{MP}$ (Table 3).

## DISCUSSION

### *Jatropha curcas* Plantation in Brazil Currently Relies on a Limited Genetic Base

This study reports the first comprehensive assessment of the genetic diversity of cultivated *J. curcas* germplasm in Brazil. To the best of our knowledge this is the most representative collection of planting stock of this species assembled in the country and possibly one of the largest worldwide. Both microsatellites and RAPD markers clearly show that a very limited genetic diversity exists in the currently available material of *J. curcas* in Brazil when compared, for example, to the levels of genetic diversity found in 307 different plant studies surveyed for which dominant (AFLP, RAPD, and ISSR) and codominant microsatellite information was available (Nybom, 2004). The average genetic similarity estimated with RAPD markers among the 192 accessions is high (0.89) when compared to another preferentially allogamous and phylogenetically close crop of the same Euphorbiaceae family such as cassava (Asante and Offei, 2003; Colombo et al., 2000). This average similarity is equivalent to that described by Ram et al. (2008) (85%) in 12 *J. curcas* accessions from India evaluated with 18 RAPD primers and by Sun et al. (2008) (92%) in 58 accessions from China using seven AFLP primer combinations. Other recent reports, sampling diversity with different types of molecular markers have also corroborated similar conclusions that in general there is limited variation in the most accessible germplasm resources of *J. curcas* in various countries (Kumar et al., 2009b; Pamidimarri et al., 2009b; Subramanyam et al., 2009). In our study only 43 unique multilocus genotypes could be distinguished out of the 192, in spite of a relatively extensive genome-wide coverage with 96 RAPD primers yielding 381 scorable RAPD amplicons, 23 of which were polymorphic (Table 1), and six selected microsatellites, one of which was polymorphic (Table 2). This is a relevant result as it demonstrates the very limited genetic base on which the expanding front of *J. curcas* plantation in Brazil currently relies.

### Genetic Diversity and Geographic Origin of *J. curcas* in Brazil Have No Connection

The UPGMA dendrogram based on coefficients of genetic similarity showed a large cluster of accessions with a very limited diversity among them in spite of their wide geographical origin, unequivocally suggesting a common origin of several accessions. Similar results with a limited number of divergent clusters were reported by Basha and Sujatha (2007) when looking at 42 Indian accessions of *J. curcas* with 83% of them clustered in only two groups. In a later study Basha and Sujatha (2009) studying 72 genotypes of *J. curcas* from different regions of the world found only two groups, one consisting of accessions from the Central American region while the other with accessions from other regions in the world. Tatikonda et al. (2009) analyzed 48 accessions of *J. curcas* collected from six states of India with a large set of 680 polymorphic AFLP markers. In spite of the much higher resolution provided by the large number of polymorphic markers they also found only five distinguishable clusters that separated at Jaccard genetic similarities between 0.55 and 0.78, a range comparable to the one found in our study.

Several accessions displayed indistinguishable multilocus marker genotypes and were grouped in specific profiles (Table 3). Every one of these groups involved accessions collected in different states spanning an extensive latitudinal range. Conversely, some accessions collected in the same or very close localities had distinguishable marker profiles and clustered in different groups. These results clearly show that even the modest genetic diversity observed in this germplasm collection has no relationship with geographic origin. This suggests that since its recent introduction, *J. curcas* germplasm has likely experienced a widespread dispersion by human intervention across distant regions and localities by seeds and possibly by vegetative propagules as well. It is interesting to note, however, that in spite of the relatively narrow genetic base evidenced by this molecular analysis, the fact that this crop has been successfully cultivated throughout Brazil suggests that it carries a high phenotypic plasticity enabling it to adapt to a wide variety of environments.

### Extensive Duplications are Present in the *J. curcas* Germplasm Collection

Given the comparative results obtained with RAPD and microsatellite markers concerning the ability to discriminate *J. curcas* accessions, our study corroborates previous reports (Sun et al., 2008; Ram et al., 2008) showing that the higher multiplexing ability of arbitrarily primed technologies such as AFLP and RAPD is more valuable than the codominance of single locus microsatellites for this particular purpose but provide less genetic information per locus due to its dominant behavior. Although codominant microsatellites in theory are more powerful for individual discrimination, in our study the six selected microsatellites selected for high polymorphism surprisingly did not yield the expected resolution power. Furthermore, even after screening a relatively large set of arbitrary sequence primers and carefully selecting repeatable RAPD markers, several accessions displayed indistinguishable multilocus profiles.

Redundancy in germplasm collections and approaches to identify duplicates using molecular markers have been

successfully reported for a number of crops (Irisha et al., 2010; Lund et al., 2003; van Treuren et al., 2001, 2004). In this study, basic principles of random match probability were used to estimate the probability of accessions with indistinguishable multilocus profiles being true duplicates. The random match probability assumes independence (linkage equilibrium) among the markers that constitute the multilocus profile, a condition that was not tested. However algorithms to obtain the maximum-likelihood estimates of the linkage disequilibria between dominant markers have been developed and could be applied to this *J. curcas* dataset (Li et al., 2007). Probabilities of duplicated accessions varied between 83 and 99.98% depending on the profile frequency, which in turn was a result of frequencies of the alternative alleles at each individual marker. Although approximate, these estimates should prove useful for the management of *J. curcas* genetic resources when prioritizing accessions to compose breeding populations or as a preliminary indication of the level of redundancy in germplasm banks. It is important to note, however, that these probabilities of being true duplicates should be taken with due caution. Accessions with indistinguishable marker profiles could contain single point mutations conferring important phenotypic diversity for adaptation. In a still relatively unknown crop such as *J. curcas*, only after a detailed phenotypic assessment could potentially redundant accessions at the molecular level be precluded. With the availability of a reference genome for *J. curcas*, coupled to next generation low coverage resequencing of a set of accessions, an effort to develop targeted genome-wide single nucleotide polymorphism (SNP) markers could help reveal functional polymorphism underlying relevant phenotypic differences found in this germplasm collection.

The state of Minas Gerais was the most intensively surveyed section of the Brazilian territory with 128 accessions in total (67%) clustered in 29 indistinguishable profiles. The eight accessions provided by the EPAMIG germplasm bank displayed six unique profiles. Only two accessions were therefore potentially duplicated, one in the set 55-Npo-MG and 58-Npo-MG (profile C) with probability of 96.89% and another one in the set 52-Npo-MG and 59-Npo-MG (profile A) with probability 86.94%. In the UFLA germplasm bank, the 21 accessions had nine unique marker profiles and a relatively large number of accessions were potential duplicates: accessions 89-Lav-MG, 93-Lav-MG, 100-Lav-MG, 101-Lav-MG, and 103-Lav-MG (profile A) with probability of 86.94% of being duplicates; accessions 91-Lav-MG, 92-Lav-MG, 96-Lav-MG, 97-Lav-MG, 99-Lav-MG, and 104-Lav-MG (profile B) with probability 83.48%; and accessions 105-Lav-MG, 106-Lav-MG, and107-Lav-MG (profile E), with probability of 95.61%. These results indicate that these two sampled germplasm repositories display overlapping accessions at the molecular level and thus contribute limited diversity to the crop. Conversely, the relatively large number of 24 accessions collected in the field that have unique molecular profiles, indicates that the sampling expeditions were fruitful in capturing additional genetic diversity, at least at the molecular level. It remains to be seen, however, if the modest genetic diversity contained in these unique accessions and displayed by likely neutral markers will in fact prove useful for breeding purposes.

## Microsatellites Reveal Recent Origin and Suggest Extensive Homozygosity in *J. curcas* Cultivated Germplasm

The microsatellite marker data revealed two important results. First, only one out of six microsatellites used showed some level of polymorphism, albeit with only two alleles, one of them very rare. These six microsatellites were carefully selected from a set of 12 developed by Pamidimarri et al. (2009a), specifically choosing those that had the highest number of alleles, high heterozygosity, and in Hardy Weinberg equilibrium in a set of 32 accessions collected in India. In that set of germplasm, these microsatellites revealed between 4 alleles (jcds10, jcps9, and jcps6) and up to nine alleles (jcds20) or 11 alleles (jcds24). No specific information was provided regarding the origin of those 32 accessions in that study. Results of the microsatellite analysis in our germplasm collection are well in line with other *J. curcas* studies that looked at cultivated germplasm. Sun et al. (2008) reported that 16 microsatellite markers yielded a monomorphic product in 58 Chinese accessions. Basha et al. (2009) found that only accessions from Mexico and El Salvador showed variation by genotyping 12 microsatellite markers. In this same study, accessions from other countries were also tested and could not be distinguished with these markers. Although it could be argued that a larger number of microsatellites with more variable simple sequence repeat (SSR) motifs need to be examined, all these results from independent studies strongly suggest that the *J. curcas* germplasm currently cultivated across the globe has such a common and recent origin that even hypervariable sequences such as microsatellites have not had time to evolve new alleles.

A second result provided by the microsatellite analysis is the very high rate of homozigosity seen in all the 192 accessions, in contrast with the preferentially outcrossing mating system reported for *J. curcas* (Luo et al., 2007). As a monoecious species, with separate male and female flowers on the same inflorescence and temporal difference in anthesis, it is expected that most loci in *J. curcas* would be heterozygous due to the preferential occurrence of outcrossing. In this germplasm collection, however, four loci were homozygous in all 192 accessions and only nine in 166 successfully typed accessions were heterozygous at locus jcps20 (Table 2). Furthermore, all the accessions were apparently equally heterozygous at marker jcps6, which

strongly suggests a duplicated microsatellite locus with homozygous genotypes at both of them. This hypothesis evidently needs verification by either linkage mapping in segregating progenies or by sequencing each allele. Similarly, Sun et al. (2008) also found homozygozity for 16 microsatellite loci analyzed in 58 accessions of *J. curcas* from China. These results taken together question the assumption of a preferential outbred mating system for cultivated *J. curcas* in spite of its typical allogamous flowering habit. These results further suggest that cultivated *J. curcas* in Brazil and possibly in China not only support selfing but possibly preferentially breed through geitonogamy, that is, by pollination of a male flower on a female flower of the same flowering plant or a nearby plant of the same genotype.

## CONCLUSIONS

This study provides a comprehensive picture of the molecular diversity of *J. curcas* cultivated germplasm in Brazil, a country that holds a leading position in ethanol biofuel crops and a promising outlook for the expansion of oil biofuel crops (Nass et al., 2007). The limited genetic diversity and considerable extent of duplications found in this germplasm collection in spite of its widespread origin and large size most likely reflects a common ancestry, drift, and intensive selection of the currently cultivated material since its introduction in Brazil. This result highlights an urgent need to enhance this germplasm resource with novel accessions, particularly from the center of origin of the species, therefore tapping into the necessary genetic diversity to drive a vigorous breeding program for the crop.

### Acknowledgments

This work was supported by competitive grants of the Brazilian Ministry of Science and Technology through FINEP and CNPq. We also acknowledge the support of EMBRAPA competitive Macroprogram grants and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais). TBR received a posdoctoral fellowship and DG and BQ research productivity fellowships from CNPq.

### References

Abdelgadir, H.A., S.D. Johnson, and J. Van Staden. 2009. Pollinator effectiveness, breeding system, and tests for inbreeding depression in the biofuel seed crop, *Jatropha curcas*. J. Hortic. Sci. Biotechnol. 84:319–324.

Asante, I.K., and S.K. Offei. 2003. RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. Euphytica 131:113–119.

Basha, S.D., G. Francis, H.P.S. Makkar, K. Becker, and M. Sujatha. 2009. A comparative study of biochemical traits and molecular markers for assessment of genetic relationships between *Jatropha curcas* L. germplasm from different countries. Plant Sci. 176:812–823.

Basha, S.D., and M. Sujatha. 2007. Inter and intra-population variability of *Jatropha curcas* (L.) characterized by RAPD and ISSR markers and development of population-specific SCAR markers. Euphytica 156:375–386.

Basha, S.D., and M. Sujatha. 2009. Genetic analysis of *Jatropha* species and interspecific hybrids of *Jatropha curcas* using nuclear and organelle specific markers. Euphytica 168:197–214.

Berchmans, H.J., and S. Hirata. 2008. Biodiesel production from crude *Jatropha curcas* L. seed oil with a high content of free fatty acids. Bioresour. Technol. 99:1716–1721.

Besnard, G., C. Breton, P. Baradat, B. Khadari, and A. Berville. 2001. Cultivar identification in olive based on RAPD markers. J. Am. Soc. Hortic. Sci. 126:668–675.

Brondani, R.P.V., and D. Grattapaglia. 2001. Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. Biotechniques 31:793.

Carels, N. 2009. *Jatropha curcas*: A Review. Adv. Bot. Res. 50:39–86.

Carvalho, C.R., W.R. Clarindo, M.M. Praca, F.S. Araujo, and N. Carels. 2008. Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. Plant Sci. 174:613–617.

Colombo, C., G. Second, and A. Charrier. 2000. Diversity within American cassava germ plasm based on RAPD markers. Genet. Mol. Biol. 23:189–199.

Doyle, J.J., and J.L. Doyle. 1990. Isolation of fresh DNA from fresh tissue. Focus 12:3–15.

Evett, I.W., and B. Weir. 1998. Interpreting DNA evidence. Sinauer Associates Inc., Sunderland, MA.

Fairless, D. 2007. Biofuel: The little shrub that could – maybe. Nature 449:652–655.

Francis, G., R. Edinger, and K. Becker. 2005. A concept for simultaneous wasteland reclamation, fuel production, and socio-economic development in degraded areas in India: Need, potential and perspectives of *Jatropha* plantations. Nat. Resour. Forum 29:12–24.

Ginwal, H.S., S.S. Phartyal, P.S. Rawat, and R.L. Srivastava. 2005. Seed source variation in morphology, germination and seedling growth of *Jatropha curcas* Linn. in central India. Silvae Genet. 54:76–80.

Gorg, R., U. Schachtschabel, E. Ritter, F. Salamini, and C. Gebhardt. 1992. Discrimination among 136 tetraploid potato varieties by fingerprints using highly polymorphic DNA Markers. Crop Sci. 32:815–819.

Grattapaglia, D., and R. Sederoff. 1994. Genetic-linkage maps of eucalyptus-grandis and eucalyptus-urophylla using a pseudo-testcross-mapping strategy and RAPD markers. Genetics 137:1121–1137.

Gupta, S., M. Srivastava, G.P. Mishra, P.K. Naik, R.S. Chauhan, S.K. Tiwari, M. Kumar, and R. Singh. 2008. Analogy of ISSR and RAPD markers for comparative analysis of genetic diversity among different *Jatropha curcas* genotypes. Afr. J. Biotechnol. 7:4230–4243.

Idury, R.M., and L.R. Cardon. 1997. A simple method for automated allele binning in microsatellite markers. Genome Res. 7:1104–1109.

Irisha, B.M., R. Goenaga, D. Zhang, R. Schnell, J.S. Brown, and J.C. Motamayor. 2010. Microsatellite fingerprinting of the USDA-ARS Tropical Agriculture Research Station Cacao (*Theobroma cacao* L.) germplasm collection. Crop Sci. 50:656–667.

Jaccard, P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. (In French.) Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579.

Kumar, R.S., K.T. Parthiban, and M.G. Rao. 2009a. Molecular characterization of *Jatropha* genetic resources through inter-simple sequence repeat (ISSR) markers. Mol. Biol. Rep. 36:1951–1956.

Kumar, R.V., Y.K. Tripathi, P. Shukla, S.P. Ahlawat, and V.K. Gupta. 2009b. Genetic diversity and relationships among germplasm of *Jatropha curcas* L. revealed by RAPDs.

Trees-Structure and Function 23:1075–1079.
Li, Y.C., Y. Li, S. Wu, K. Han, Z.J. Wang, W. Hou, Y.R. Zeng, and R.L. Wu. 2007. Estimation of multilocus linkage disequilibria in diploid populations with dominant markers. Genetics 176:1811–1821.
Liu, H.F., Y.F. Deng, and J.P. Liao. 2008. Floral organogenesis of three species of *Jatropha* (Euphorbiaceae). J. Syst. Evol. 46:53–61.
Lund, B., R. Ortiz, I.M. Skovgaard, R. Waugh, and S.B. Andersen. 2003. Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor. Appl. Genet. 106:1129–1138.
Luo, C.-w., K. Li, Y. Chen, and Y.-y. Sun. 2007. Floral display and breeding system of *Jatropha curcas* L. For. Stud. China 9:114–119.
Modi, M.K., J.R.C. Reddy, B.V.S.K. Rao, and R.B.N. Prasad. 2007. Lipase-mediated conversion of vegetable oils into biodiesel using ethyl acetate as acyl acceptor. Bioresour. Technol. 98:1260–1264.
Nass, L.L., P.A.A. Pereira, and D. Ellis. 2007. Biofuels in Brazil: An overview. Crop Sci. 47:2228–2237.
Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol. Ecol. 13:1143–1155.
Openshaw, K. 2000. A review of *Jatropha curcas*: An oil plant of unfulfilled promise. Biomass Bioenergy 19:1–15.
Pamidimarri, D.V.N.S., R. Sinha, P. Kothari, and M.P. Reddy. 2009a. Isolation of novel microsatellites from *Jatropha curcas* L. and their cross-species amplification. Mol. Ecol. Resour. 9:431–433.
Pamidimarri, D.V.N.S., S. Singh, S.G. Mastan, J. Patel, and M.P. Reddy. 2009b. Molecular characterization and identification of markers for toxic and non-toxic varieties of *Jatropha curcas* L. using RAPD, AFLP and SSR markers. Mol. Biol. Rep. 36:1357–1364.
Prevost, A., and M.J. Wilkinson. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor. Appl. Genet. 98:107–112.
Ram, S.G., K.T. Parthiban, R.S. Kumar, V. Thiruvengadam, and M. Paramathma. 2008. Genetic diversity among *Jatropha* species as revealed by RAPD markers. Genet. Resour. Crop Evol. 55:803–809.
Reis, A.M.M., and D. Grattapaglia. 2004. RAPD variation in a germplasm collection of *Myracrodruon urundeuva* (Anacardiaceae), an endangered tropical tree: Recommendations for conservation. Genet. Resour. Crop Evol. 51:529–538.
Rohlf, F.J. 2002. Numerical taxonomy and multivariate analysis system, ver. 2.1. Exeter Ltd., Setauket, NY.
Saturnino, H.M., D.D. Pacheco, J. Kakida, N. Tominaga, and N.P. Gonçalves. 2005. Cultura do pinhão-manso (*Jatropha curcas* L.). (In Spanish.) Informe Agropecuário 26:44–78.
Sneath, P.H.A., and R.R. Sokal. 1973. Numerical taxonomy. Freeman Press, San Francisco, CA.
Subramanyam, K., D. Muralidhararao, and N. Devanna. 2009. Genetic diversity assessment of wild and cultivated varieties of *Jatropha curcas* (L.) in India by RAPD analysis. Afr. J. Biotechnol. 8:1900–1910.
Sun, Q.B., L.F. Li, Y. Li, G.J. Wu, and X.J. Ge. 2008. SSR and AFLP markers reveal low genetic diversity in the biofuel plant *Jatropha curcas* in China. Crop Sci. 48:1865–1871.
Tatikonda, L., S.P. Wani, S. Kannan, N. Beerelli, T.K. Sreedevi, D.A. Hoisington, P. Devi, and R.K. Varshney. 2009. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L., a biofuel plant. Plant Sci. 176:505–513.
van Treuren, R., L.J.M. van Soest, and T.J.L. van Hintum. 2001. Marker-assisted rationalisation of genetic resource collections: A case study in flax using AFLPs. Theor. Appl. Genet. 103:144–152.
van Treuren, R., A. Magda, R. Hoekstra, and T.J.L. van Hintum. 2004. Genetic and economic aspects of marker-assisted reduction of redundancy from a wild potato germplasm collection. Genet. Resour. Crop Evol. 51:277–290.